

Messung des mathematischen Textverständnisses von Altenpflegeschülern zu Beginn der Pflegeausbildung

Eine empirische Untersuchung an privaten und öffentlichen Altenpflegeschulen

Inaugural-Dissertation

zur Erlangung des
Doktorgrades der Pflegewissenschaft
(Dr. rer. cur.)

an der Pflegewissenschaftlichen Fakultät der Philosophisch-
Theologischen Hochschule Vallendar

vorgelegt von:
Thomas Knappich

Erstgutachter:
Prof. Dr. Albert Brühl

Zweitgutachter:
Prof. Dr. Martin Stein

vorgelegt im Januar 2017

Danksagung

Bedanken möchte ich mich ganz herzlich bei meinem Doktorvater Herrn Prof. Dr. Albert Brühl für seine Hilfe und motivierende Unterstützung. Mit seiner freundlichen und stets wertschätzenden Art begleitete er mich während der Erstellung meiner Dissertation. Der intensive fachliche Austausch war sehr hilfreich. Darüber hinaus möchte ich mich bei Herrn Prof. Dr. Martin Stein für seine fachliche Unterstützung bedanken.

Gleichermaßen gilt mein Dank den Mitarbeitenden der Pflegeschulen, die bereit dazu waren, die erforderlichen Daten zu erheben. Darüber hinaus möchte ich mich bei meinen – auch ehemaligen – Kommilitoninnen und Kommilitonen des Promotionskolloquiums an der Philosophisch Theologischen Hochschule in Vallendar für die anregenden Diskussionen und die wegweisenden Anregungen bedanken.

Ein ganz großes Dankeschön geht an meine Familie, insbesondere an meine Frau Miriam, die mich immer unterstützt und mir die nötigen Freiräume verschafft hat.

Für eine bessere Lesbarkeit der vorliegenden Arbeit beschränke ich mich sprachlich auf die männliche Form von Personen. Dies bedeutet keinesfalls eine Benachteiligung weiblicher Personen. Wenn beispielsweise vom Altenpflegeschüler gesprochen wird, so beinhaltet dies selbstverständlich auch die Altenpflegeschülerin.

Abstract

Hintergrund:

Basiskompetenzen respektive Kulturtechniken, zu welchen auch das mathematische Textverständnis gehört, sind für eine erfolgreiche Gestaltung eines Ausbildungsverlaufes, insbesondere in der Altenpflegeausbildung, von großer Bedeutung. Aufgrund der Tatsache, dass nach eingehender Sichtung der Literatur und der Studienlage auf keine Vorarbeiten zurückgegriffen werden konnte, die sich primär um die Zielgruppe der Pflege beschäftigten, wurde der Test zur Erfassung des mathematischen Textverständnisses (Jordan & Stein 2011c) gewählt. Es stellt sich die Frage, ob mit den vorliegenden Testverfahren das mathematische Textverständnis von Altenpflegeschülern valide erfasst werden kann.

Methode:

Für diese Untersuchung wurden insgesamt 664 Altenpflegeschüler aus 17 bundesweit angesiedelten privaten und öffentlichen Altenpflegeschulen befragt. Unter Anwendung der Probabilistischen Testtheorie wurde geprüft, ob das Rasch-Modell für die dichotomen Daten gilt, welche mit dem TeMaTex und dem CareMaTex generiert wurden. Darüber hinaus wurde überprüft, ob die Summenwerte aus den dichotomen Daten aussagekräftig in Kompetenzstufen beziehungsweise Noten transformiert werden dürfen.

Ergebnisse:

15 der ursprünglich 18 Items des TeMaTex konnten an den Kontext der Altenpflege angepasst und in Form eines neuen validen Tests zur Erfassung des mathematischen Textverständnisses (CareMaTex) dargestellt werden. Vor diesem Hintergrund ließen sich Aussagen zum Niveau der Altenpflegeschüler hinsichtlich des mathematischen Textverständnisses ableiten. Insgesamt steigen die Hauptschüler und Realschüler im Vergleich zu den Schülern mit Fachhochschulreife und Abitur mit einem geringeren Niveau in die Altenpflegeausbildung ein.

Schlussfolgerungen:

Es wird nicht empfohlen, im Rahmen eines Bewerberauswahlverfahrens oder für die Einzelfalldiagnostik in der Altenpflegeausbildung den TeMaTex zu verwenden. Der in dieser Arbeit entwickelte CareMaTex stellt eine valide Alternative zum TeMaTex dar. Es sei festgehalten, dass die Diagnostik solcher Defizite in der Berufsausbildung nicht allein zur Lösung der Problematik beiträgt. Es müssen daran anschließend Fördermaßnahmen entwickelt werden, die dieser Problematik entgegenwirken können. Dies konnte in der vorliegenden Arbeit nicht bewerkstelligt werden und wird daher als weiterer Forschungsbedarf festgestellt.

Inhaltsverzeichnis

1	Einleitung	8
1.1	<i>Problembeschreibung und Hintergrund</i>	<i>8</i>
1.2	<i>Forschungsfragen und Hypothesen</i>	<i>11</i>
1.3	<i>Ziele und Absichten</i>	<i>13</i>
I.	Theoretischer Teil.....	15
2	Rechtliche Rahmenbedingungen des Altenpflegeberufs	15
2.1	<i>Vorgaben des novellierten Altenpflegegesetzes.....</i>	<i>15</i>
2.2	<i>Auswirkungen der Gesetzesänderung</i>	<i>18</i>
2.3	<i>Kompetenzfeststellungsverfahren im Rahmen der Novellierung des Altenpflegegesetzes.....</i>	<i>18</i>
2.4	<i>Zusammenfassende Betrachtung des zweiten Kapitels</i>	<i>21</i>
3	Mathematisches Berufsprofil der Altenpflege	22
3.1	<i>Analyse mathematischer Anforderungen im curricularen Bereich</i>	<i>23</i>
3.2	<i>Analyse mathematischer Anforderungen anhand ausgewählter Aufgaben aus der theoretischen Ausbildung</i>	<i>27</i>
3.3	<i>Mathematische Anforderungsbereiche in der Ausbildung.....</i>	<i>30</i>
3.4	<i>Zusammenfassende Betrachtung des dritten Kapitels</i>	<i>35</i>
4	Mathematisches Textverständnis.....	37
4.1	<i>Begriffsbestimmung</i>	<i>37</i>
4.2	<i>Textaufgaben als Aufgabentyp zur Erfassung von mathematischem Textverständnis.....</i>	<i>41</i>
4.3	<i>Strategien zum Lösen und Verstehen von Textaufgaben</i>	<i>43</i>
4.4	<i>Zusammenfassende Betrachtung des vierten Kapitels.....</i>	<i>47</i>
II.	Methodischer Teil	49
5	Methoden und Vorgehensweisen der Studie.....	49
5.1	<i>Ethische Überlegungen und Datenschutz</i>	<i>49</i>
5.2	<i>Stichprobengrößen</i>	<i>50</i>
5.3	<i>Beschreibung der Datenerhebungen.....</i>	<i>51</i>
5.4	<i>Durchführung der Datenauswertungen</i>	<i>52</i>
5.5	<i>Expertenbefragung</i>	<i>54</i>
5.6	<i>Bestimmung der Gütekriterien eines Testverfahrens.....</i>	<i>57</i>
5.7	<i>Diskussion der Methoden.....</i>	<i>62</i>
5.8	<i>Zusammenfassende Betrachtung des fünften Kapitels.....</i>	<i>65</i>
6	Testtheorie und Testmodelle	66
6.1	<i>Probabilistische Testtheorie</i>	<i>66</i>
6.1.1	<i>Grundlagen des dichotomen 1-PL-Modells nach Rasch</i>	<i>67</i>

6.1.2	Parameterschätzung im Rasch-Modell	69
6.1.3	Parameterschätzung bei missing data	71
6.1.4	Differenz zwischen Personen- und Itemparameter	72
6.1.5	Modellgeltungstests bei dichotomem 1-PL-Rasch-Modell	73
6.1.5.1	Grafischer Modellgeltungstest	73
6.1.5.2	Wald-Test	74
6.1.5.3	Tests zur Überprüfung der Signifikanz	74
6.1.5.3.1	Likelihood-Quotienten-Test	75
6.1.5.3.2	Pearson χ^2 Test	75
6.1.5.3.3	Bootstrap-Verfahren	75
6.1.5.3.4	Lokale Modellgeltungstests	76
6.1.6	Differential Item Functioning	78
6.2	Zusammenfassende Betrachtung des sechsten Kapitels	78
7	Test zur Erfassung des mathematischen Textverständnisses (TeMaTex)	80
7.1	Beschreibung des Tests und der Aufgaben	80
7.2	Gütekriterien des Tests	90
7.2.1	Objektivität	90
7.2.2	Reliabilität	91
7.2.3	Validität	92
7.3	Zusammenfassende Betrachtung des siebten Kapitels	92
III.	Empirischer Teil 1	95
8	Statistische Analysen in Bezug auf den TeMaTex	95
8.1	Deskriptive Analyse	95
8.2	Inferenzstatistische Analyse	98
8.2.1	Ergebnisse der Modellentwicklung mit dem Rasch-Modell	98
8.2.1.1	Rasch-Analyse mit RUMM 2030	99
8.2.1.2	Rasch-Analyse mit WINMIRA	105
8.2.1.3	Rasch-Analyse mit eRm	109
8.3	Zusammenfassende Betrachtung des achten Kapitels	114
IV.	Empirischer Teil 2	115
9	Expertenrating in Bezug auf die Items des CareMaTex	115
9.1	Expertenbefragung	115
9.1.1	Antwortverhalten der Experten	115
9.1.2	Itemanalyse aufgrund des Expertenurteils	116
9.1.3	Ermittlung der Urteilsübereinstimmung	118
9.2	Zusammenfassende Betrachtung des neunten Kapitels	120
10	Statistische Analysen in Bezug auf den CareMaTex	121
10.1	Beschreibende statistische Analyse	121

10.2	<i>Schließende statistische Analyse</i>	124
10.2.1	<i>Ergebnisse aus der Modellentwicklung anhand des Rasch-Modells</i>	124
10.2.1.1	<i>Prüfung der Modellgeltung mit RUMM 2030</i>	125
10.2.1.2	<i>Prüfung der Modellgeltung mit WINMIRA</i>	128
10.2.1.3	<i>Prüfung der Modellgeltung mit eRm</i>	132
10.3	<i>Zusammenfassende Betrachtung des zehnten Kapitels</i>	136
11	Beschreibung des CareMaTex	137
11.1	<i>Darlegung der Items des CareMaTex</i>	137
11.2	<i>Gütekriterien des CareMaTex</i>	145
11.2.1	<i>Kriterium der Objektivität</i>	145
11.2.2	<i>Kriterium der Reliabilität</i>	146
11.2.3	<i>Kriterium der Validität</i>	146
11.3	<i>Zusammenfassende Betrachtung des elften Kapitels</i>	146
V.	Diskussion	147
12	Diskussion der Ergebnisse	147
12.1	<i>Diskussion der Ergebnisse der statistischen Analysen in Bezug auf den TeMaTex</i>	147
12.2	<i>Diskussion der Ergebnisse der statistischen Analysen in Bezug auf den CareMaTex</i>	157
12.3	<i>Diskussion zu den Forschungsfragen und Hypothesen</i>	165
13	Fazit und Ausblick	170
14	Literatur	174
15	Verzeichnis der Abkürzungen	180
16	Tabellenverzeichnis	182
17	Abbildungsverzeichnis	186
18	Anhang	187

1 Einleitung

Im Rahmen der Einleitung wird sowie das Problem beschrieben als auch der Hintergrund der vorliegenden Arbeit erläutert. Darüber hinaus werden die Forschungsfragen und die daraus abgeleiteten Hypothesen beschrieben. Abschließend findet eine Darstellung der Zielsetzung und Absicht der Arbeit statt.

1.1 Problembeschreibung und Hintergrund

Mit dem von der Bundesregierung initiierten Gesetz zur Stärkung der beruflichen Aus- und Weiterbildung in der Altenpflege vom 13. März 2013 (BGBl. I, S. 446) im Rahmen der Ausbildungs- und Qualifizierungsoffensive in der Altenpflege haben sich unter anderem die Zugangsvoraussetzungen für die Altenpflegeausbildung verändert. Das Gesetz sieht eine Erschließung des Nachqualifizierungspotentials in der Altenpflege vor, somit sollen geeignete qualifizierte Nachwuchskräfte für die Altenpflegeausbildung rekrutiert werden. Bei Vorliegen entsprechender Qualifikationen ist sogar eine Verkürzung der Ausbildungszeit möglich. Die Entscheidung, ob und in welchem Umfang die Ausbildungszeit eines Bewerbers verkürzt werden kann, soll mittels einer Kompetenzmessung durch die entsprechenden Schul- und Landesbehörden festgestellt werden (vgl. BGBl. I, S. 446). Nur ist noch nicht eindeutig ersichtlich, wie eine solche Kompetenzfeststellung auszusehen hat und welche Kompetenzen gemessen werden sollen.

Die Novellierung des Altenpflegegesetzes kann als ein Zeichen dafür gesehen werden, dass sich das Arbeitsumfeld in der Altenpflege in den letzten zehn Jahren stark verändert hat. Gartz spricht von einem Wandel der Arbeitswelt, der neue betriebliche Anforderungen an die (angehenden) Altenpflegefachkräfte mit sich bringt (vgl. Gartz et al. 1999, S. 10).

„Mit der Entwicklung von der Produktions- hin zur Dienstleistungs-, Freizeit-, Informations- und Bildungsgesellschaft vollzieht sich ein Prozeß, an dessen Ende eine grundlegend veränderte Arbeitswelt stehen wird. Einfache und monoton wiederkehrende Arbeitsvorgänge verlieren an Bedeutung gegenüber der Verarbeitung und Anwendung von Informationen und theoretischem Wissen. Ebenso hat der technische Wandel ein Ende des bislang stattfindenden Arbeitsteilungsprozesses

mit sich gebracht. Auch die Inhalte der Arbeit, ihre Organisationsformen und ihre gesellschaftliche Bewertung verändern sich tiefgreifend. Damit geht ein markanter Wandel der Qualifikationsanforderungen an die Mitarbeiter und folglich an die gegenwärtigen und künftigen Auszubildenden einher.“ (Gartz et al. 1999, S. 10)

Seit dem Inkrafttreten des neuen Altenpflegegesetzes (2003) und der Ausbildungs- und Prüfungsverordnung (2002) wird versucht, den Anforderungen der sich stetig verändernden Arbeitswelt durch einen lernfeld- und handlungsorientierten Unterricht in der Altenpflegeausbildung gerecht zu werden. Im Rahmen eines solchen Unterrichts sollen die Auszubildenden komplexe Lernsituationen aus der realen Lebens- und Arbeitswelt bearbeiten. Hierfür werden bei den Lernenden unterschiedliche Fertigkeiten vorausgesetzt. Von den Auszubildenden wird sogar explizit erwartet, dass sie die Fähigkeit mitbringen, selbständig berufsbezogene Texte unterschiedlicher Themen und Lernbereiche erschließen zu können. „Die Arbeit mit Texten ist in der Berufsschule ein elementarer Gegenstand.“ (Becker-Mrotzek & Kusch 2007, S. 33) Die Auszubildenden in den Berufsschulen und Berufsfachschulen der Altenpflege arbeiten überwiegend mit Fachtexten, die aus einer Kombination von Fließtext, Diagrammen und Tabellen bestehen. Die Anforderung an die Lernenden besteht darin, diese Textsorten lesen und verstehen respektive verarbeiten zu können. Aber nicht jeder Auszubildende fühlt sich von den allgemein bildenden Schulen gut auf die Anforderungen in der Berufsschule hinsichtlich der Textverwertung vorbereitet (vgl. Pospiech & Bitterlich 2007, S. 19 ff.).

Der Bildungsbericht 2012 verdeutlicht die brisante Situation in der beruflichen (Aus-)Bildung anhand konkreter Zahlen. 2011 wurden ungefähr 600.000 Ausbildungsstellen angeboten. Die Nachfrage lag hingegen bei circa 580.000 Ausbildungsstellen. Insgesamt wurden etwa 570.000 Ausbildungsverträge neu abgeschlossen (vgl. Deutsches Institut für Internationale Pädagogische Forschung 2012, S. 107).

Als bedingt ausbildungsfähig werden jene Auszubildende bezeichnet, die den betrieblichen Anforderungen aufgrund mangelhafter Handhabung der Kulturtechniken wie Lesen, Schreiben und Rechnen nicht gerecht werden

können. Branchenübergreifend werden von den Auszubildenden solide Kenntnisse in Bezug auf die Kulturtechniken gefordert (vgl. Gartz et al. 1999, S. 132 f.). Denn ohne ausreichende Kulturtechniken „keine qualifizierte berufliche Ausbildung, ohne qualifizierte berufliche Ausbildung keine Integration in die Arbeitswelt und ohne Letzteres auch keine Hoffnung auf gesellschaftliche Integration und auf ein Leben mit Perspektive.“ (Grundmann 2007, S. 72). Der prozentuale Anteil der nicht oder nur bedingt ausbildungsfähigen Schulabgänger liegt zwischen 25 % und 30 % (vgl. Grundmann 2007, S. 74).

Die *International Conference on Adult Education* (UIE) konstatiert, dass jeder Mensch das Recht auf Grundbildung als den Erwerb von Basiskompetenzen wie Lesen und Rechnen besitze (vgl. UIE 1997, S. 4).

„Grundbildung für alle bedeutet, dass Menschen ungeachtet ihres Alters die Möglichkeit haben, als Einzelne oder in der Gemeinschaft ihr Potential zu entfalten. Sie ist nicht nur ein Recht, sondern auch eine Pflicht und eine Verantwortung gegenüber anderen und der Gesellschaft als Ganzem. Es ist wichtig, dass die Anerkennung des Rechts auf lebenslanges Lernen von Maßnahmen flankiert wird, die die Voraussetzungen für die Ausübung dieses Rechts schaffen. Mit den Herausforderungen des 21. Jahrhunderts können Regierungen, Organisationen oder Institutionen allein nicht fertig werden; sie brauchen die Energie, die Phantasie und schöpferische Kraft der Menschen und ihre umfassende, uneingeschränkte und tatkräftige Mitwirkung in allen Lebensbereichen. Lernen im Jugend- und Erwachsenenalter ist eines der wichtigsten Mittel, um Kreativität und Produktivität im weitesten Sinne erheblich zu verstärken, und dies wiederum ist unverzichtbar, wenn wir die komplexen, miteinander in Wechselbeziehung stehenden Probleme einer Welt lösen wollen, die einem immer rascheren Wandel, zunehmender Komplexität und einem wachsenden Risiko ausgesetzt ist“ (UIE 1997, S. 4).

Solche *Basiskompetenzen* umfassen also Fähigkeiten, die notwendig sind, um eine befriedigende Lebensführung im privaten und beruflichen Umfeld zu erreichen und am gesellschaftlichen Leben sinnvoll teilnehmen zu können (vgl. Artelt et al. 2001, S. 78).

Basierend auf diesen Erkenntnissen wird im Rahmen der vorliegenden Arbeit versucht, mithilfe des Tests zur Erfassung des mathematischen Textverständnisses (TeMaTex) zwei Teilbereiche der Kulturtechniken zu erfassen.

sen. Der TeMaTex erfasst das Textverständnis und zugleich mathematische Grundkenntnisse (vgl. Jordan & Stein 2011b). Eine ausführliche Darstellung des Testverfahrens erfolgt in Kapitel 7. In Kapitel 3 wird ausführlich erläutert, weshalb das Textverständnis und mathematische Grundkenntnisse eine hohe Relevanz in der Altenpflegeausbildung haben.

1.2 Forschungsfragen und Hypothesen

Aus den referierten Ausführungen lassen sich folgende zentrale Fragestellungen ableiten, die beantwortet werden sollen:

- 1) Kann mit den Items des Tests zur Erfassung des mathematischen Textverständnisses (TeMaTex) das mathematische Textverständnis von Altenpflegeschülern zu Beginn der Ausbildung gemessen werden?
- 2) Dürfen die dichotomen Bewertungen der Items des TeMaTex intervallskaliert behandelt, aufsummiert und in Notenwerte sowie in Kompetenzstufen transformiert werden?
- 3) Bilden die Items des TeMaTex das mathematische Textverständnis von Altenpflegeschülern zu Ausbildungsbeginn gesichert ab?
- 4) Lassen sich erste Aussagen treffen, auf welchem Niveau im mathematischen Textverständnis die Altenpflegeschüler in die dreijährige Altenpflegeausbildung einsteigen?
- 5) Ist es möglich, die Items des TeMaTex an den Kontext der Altenpflege anzupassen, um ein konstruktvalides Testinstrument zur Erfassung des mathematischen Textverständnisses in der Altenpflege (CareMaTex) zu konstruieren?

Aus den formulierten wissenschaftlichen Fragestellungen 1 bis 3 lassen sich folgenden Hypothesen herleiten und mit Signifikanztests prüfen:

Hypothesen zur Forschungsfrage 1:

H₀: Die Items des Tests zur Erfassung des mathematischen Textverständnisses stellen das mathematische Textverständnis von Altenpflegeschülern

zu Beginn der Ausbildung inhaltsvalide dar. Sie bilden das theoretische Konstrukt des mathematischen Textverständnisses ab.

H₁: Die Items des Tests zur Erfassung des mathematischen Textverständnisses stellen das mathematische Textverständnis von Altenpflegeschülern zu Beginn der Ausbildung nicht inhaltsvalide dar. Sie bilden das theoretische Konstrukt des mathematischen Textverständnisses nicht ab.

Hypothesen zur Forschungsfrage 2:

H₀: Die dichotomen Items des TeMaTex können intervallskaliert behandelt werden. Einzelne Bewertungen von Altenpflegeschülern dürfen aufsummiert und in Notenwerte beziehungsweise Kompetenzstufen transformiert werden.

H₁: Die dichotomen Items des TeMaTex können nicht intervallskaliert behandelt werden. Einzelne Bewertungen von Altenpflegeschülern dürfen nicht aufsummiert und in Notenwerte beziehungsweise Kompetenzstufen transformiert werden.

Hypothesen zur Forschungsfrage 3:

H₀: Nach der in dieser Arbeit durchgeführten Untersuchung bilden die Items des TeMaTex das mathematische Textverständnis von Altenpflegeschülern zu Ausbildungsbeginn konstruktvalide ab.

H₁: Nach der in dieser Arbeit durchgeführten Untersuchung bilden die Items des TeMaTex das mathematische Textverständnis von Altenpflegeschülern zu Ausbildungsbeginn nicht konstruktvalide ab.

Überprüft werden die statistischen Hypothesen mit den erfassten Leistungen der Auszubildenden in den einzelnen Items des TeMaTex, um Aussagen hinsichtlich der Gesamtmodellgeltung des Testverfahrens treffen zu können. In dieser Arbeit werden zur Berechnung der Modellgeltung verschiedene Statistik-Programme verwendet, die unterschiedliche Algorithmen zur Berechnung der Itemparameter verwenden. Daher ist die Signifikanzaussage davon abhängig, mit welchem Programm die Parameter und

die Modellgeltung berechnet werden. Hierauf wird in Kapitel 5.4 näher eingegangen.

1.3 Ziele und Absichten

Die Ergebnisse der Schulleistungsstudien – wie zum Beispiel PISA – zeigen, wie wichtig es ist, nicht nur die Basiskompetenzen von Schülern der allgemeinbildenden Schulen zu fördern, sondern auch die der Schüler in den Berufs(fach)schulen (vgl. Knappich 2012, S. 12). In diesem Zusammenhang ist es das primäre Ziel dieses Forschungsvorhabens, das Diagnoseinstrument TeMaTex, welches das vorhandene mathematische Textverständnis bei Altenpflegeschülern zu Ausbildungsbeginn erfassen soll, auf seine Validität hin zu überprüfen sowie gegebenenfalls erste Aussagen über die Ausprägung des mathematischen Textverständnisses von Altenpflegeschülern zu erarbeiten. Des Weiteren soll untersucht werden, ob die Items des TeMaTex an den Kontext der Altenpflege angepasst werden können, um daraus ein valides Testinstrument (CareMaTex) zu entwickeln.

Darüber hinaus wird sich zu Beginn mit den Rahmenbedingungen der Kompetenzmessung auseinandergesetzt, indem insbesondere auf die rechtlichen Rahmenbedingungen in Bezug auf Kompetenzbilanzierungsverfahren in der Altenpflegeausbildung eingegangen wird. Zudem wird die Relevanz des mathematischen Textverständnisses anhand eines mathematischen Berufsprofils herausgearbeitet und dargelegt. Im Anschluss wird das Konstrukt des mathematischen Textverständnisses definiert und unter anderem Möglichkeiten beschrieben, wie das Konstrukt zu messen ist. Danach erfolgt die Darlegung des methodischen Teils der Arbeit, zu dem die Beschreibung der methodischen Vorgehensweise und eine Erläuterung der Probabilistischen Testtheorie zählen. Hierbei wird unter anderem eine Übersicht über die angewendeten Testmodelle gegeben. Im methodischen Teil der Arbeit wird auch der Test zur Erfassung des mathematischen Textverständnisses (TeMaTex) erläutert.

Anschließend werden die Ergebnisse der Untersuchung aufgezeigt. Dies beinhaltet die Darlegung der deskriptiven Analyse und eine Beschreibung

der durchgeführten Rasch-Analysen. Im Rahmen des Diskussionsteils werden schließlich im ersten Schritt die Ergebnisse der Untersuchung interpretiert und diskutiert. Zum Abschluss werden noch einmal die Forschungsfragen und Hypothesen der Arbeit aufgegriffen. Ein Fazit schließt die Arbeit ab.

I.Theoretischer Teil

Im theoretischen Teil der Arbeit werden zunächst in Kapitel 2 die rechtlichen Rahmenbedingungen der Altenpflege und Altenpflegeausbildung in Deutschland beleuchtet, bevor dann in Kapitel 3 näher auf das mathematische Berufsprofil der Altenpflege eingegangen wird. In Kapitel 4 wird unter anderem das Konstrukt des mathematischen Textverständnisses erläutert.

2 Rechtliche Rahmenbedingungen des Altenpflegeberufs

In Kapitel 2.1 werden die Vorgaben des novellierten Altenpflegegesetzes vorgestellt und in Kapitel 2.2 die Auswirkungen der Gesetzesänderung aufgezeigt. Kapitel 2.3 umfasst die Darstellung der ersten modellhaften Kompetenzfeststellungsverfahren im Rahmen der Altenpflegeausbildung seit der Gesetzesnovellierung. Abschließend findet in Kapitel 2.4 eine Zusammenfassung statt.

2.1 Vorgaben des novellierten Altenpflegegesetzes

Die Ausbildung in der Altenpflege wird durch das bundeseinheitliche Altenpflegegesetz geregelt. Die pädagogische Orientierung an der Kompetenzvermittlung in der Altenpflegeausbildung entspricht dem geläufigen berufspädagogischen Ansatz. Die rechtlichen Rahmenbedingungen in der Altenpflege stellen in der vorliegenden Arbeit die Ausgangslage dar, aus der sich die Relevanz des Themas und der Untersuchung ableiten lässt.

Die Altenpflegeausbildung in Deutschland basiert auf dem Gesetz über die Berufe in der Altenpflege (AltPflG, in der Fassung der Bekanntmachung vom 25. August 2003, BGBl. I S. 1690, zuletzt geändert durch Art. 1 des Gesetzes vom 13. März 2013, BGBl. I S. 446) und der Altenpflege-Ausbildungs- und Prüfungsverordnung (AltPflAPrV vom 26. Nov. 2002, BGBl. I S. 4418, zuletzt geändert durch Art. 38 des Gesetzes vom 6. Dezember 2011, BGBl. I S. 2515), welche seit dem 25. August 2003 rechtsverbindlich sind. Das Ausbildungsziel ist in § 3 AltPflG so definiert:

„(1) Die Ausbildung in der Altenpflege soll die Kenntnisse, Fähigkeiten und Fertigkeiten vermitteln, die zur selbständigen und eigenverantwortlichen Pflege einschließlich der Beratung, Begleitung und Betreuung alter Menschen erforderlich sind. Dies umfasst insbesondere:

1. die sach- und fachkundige, den allgemein anerkannten pflegewissenschaftlichen, insbesondere den medizinisch-pflegerischen Erkenntnissen entsprechende, umfassende und geplante Pflege,
2. die Mitwirkung bei der Behandlung kranker alter Menschen einschließlich der Ausführung ärztlicher Verordnungen,
3. die Erhaltung und Wiederherstellung individueller Fähigkeiten im Rahmen geriatrischer und gerontopsychiatrischer Rehabilitationskonzepte,
4. die Mitwirkung an qualitätssichernden Maßnahmen in der Pflege, der Betreuung und der Behandlung,
5. die Gesundheitsvorsorge einschließlich der Ernährungsberatung,
6. die umfassende Begleitung Sterbender,
7. die Anleitung, Beratung und Unterstützung von Pflegekräften, die nicht Pflegefachkräfte sind,
8. die Betreuung und Beratung alter Menschen in ihren persönlichen und sozialen Angelegenheiten,
9. die Hilfe zur Erhaltung und Aktivierung der eigenständigen Lebensführung einschließlich der Förderung sozialer Kontakte und
10. die Anregung und Begleitung von Familien- und Nachbarschaftshilfe und die Beratung pflegender Angehöriger.

Darüber hinaus soll die Ausbildung dazu befähigen, mit anderen in der Altenpflege tätigen Personen zusammenzuarbeiten und diejenigen Verwaltungsarbeiten zu erledigen, die in unmittelbarem Zusammenhang mit den Aufgaben in der Altenpflege stehen.

(2) Soweit in Modellvorhaben nach § 4 Abs. 7 erweiterte Kompetenzen zur Ausübung heilkundlicher Tätigkeiten erprobt werden, hat sich die Ausbildung auch auf die Befähigung zur Ausübung der Tätigkeiten zu erstrecken, für die das Modellvorhaben qualifizieren soll. Das Nähere regeln die Lehrpläne der Altenpflegesschulen und die Ausbildungspläne der Träger der praktischen Ausbildung.“

Die einzelnen Lernbereiche respektive die untergliederten Lernfelder stehen in engem Bezug mit dem Ausbildungsziel. Die Themengebiete innerhalb eines Lernfeldes geben Auskunft, welche Kompetenzen der Auszubildende innerhalb der Ausbildung erwerben soll (vgl. AltPflG, in der Fassung der Bekanntmachung vom 25. August 2003, BGBl. I S. 1690, zuletzt geändert durch Art. 1 des Gesetzes vom 13. März 2013, BGBl. I S. 446 f.). Mit den §§ 6 und 7 AltPflG werden die Zugangsvoraussetzungen zur Ausbildung sowie die Möglichkeiten zur Ausbildungsverkürzung geregelt. In § 6 heißt es:

„Voraussetzung für den Zugang zur Ausbildung ist, dass die Bewerberin oder der Bewerber nicht in gesundheitlicher Hinsicht zur Ausübung des Berufes ungeeignet ist sowie

1. der Realschulabschluss oder ein anderer als gleichwertig anerkannter Bildungsabschluss oder eine andere abgeschlossene zehnjährige Schulbildung, die den Hauptschulabschluss erweitert, oder
2. der Hauptschulabschluss oder ein als gleichwertig anerkannter Bildungsabschluss, sofern eine erfolgreich abgeschlossene, mindestens zweijährige Berufsausbildung oder die Erlaubnis als Krankenpflegehelferin oder Krankenpflegehelfer oder eine landesrechtlich geregelte, erfolgreich abgeschlossene Ausbildung von mindestens einjähriger Dauer in der Altenpflegehilfe oder Krankenpflegehilfe nachgewiesen wird, oder
3. eine andere abgeschlossene zehnjährige allgemeine Schulbildung.“

Nach § 7 kann die dreijährige Altenpflegeausbildung unter Umständen verkürzt werden. Der Paragraph gibt den Ablauf und die Voraussetzungen für eine Verkürzung wieder. Er wurde durch Art. 1 des Gesetzes zur Stärkung der beruflichen Aus- und Weiterbildung in der Altenpflege vom 13. März 2013, BGBl. I S. 446, geändert. Er lautet nun wie folgt:

„(1) Auf Antrag kann die Dauer der Ausbildung nach § 4 Abs. 1 verkürzt werden:

1. für Altenpflegehelferinnen, Altenpflegehelfer, Krankenpflegehelferinnen, Krankenpflegehelfer, Heilerziehungspflegehelferinnen, Heilerziehungspflegehelfer, Heilerziehungshelferinnen und Heilerziehungshelfer um bis zu einem Jahr.
2. für Krankenschwestern, Krankenpfleger, Kinderkrankenschwestern, Kinderkrankenpfleger, Heilerziehungspflegerinnen und Heilerziehungspfleger mit dreijähriger Ausbildung um bis zu zwei Jahre,

(2) Auf Antrag kann die Dauer der Ausbildung nach § 4 Abs. 1 im Umfang der fachlichen Gleichwertigkeit um bis zu zwei Jahre verkürzt werden, wenn eine andere abgeschlossene Berufsausbildung nachgewiesen wird.

(3) Auf Antrag ist bei Teilnahme an einer beruflichen Weiterbildung nach dem Dritten Buch Sozialgesetzbuch für Personen gemäß Absatz 1 Nummer 2, die einschließlich der Ausbildung in einem Umfang, der einer Vollzeitbeschäftigung von mindestens zwei Jahren entspricht, in einer Pflegeeinrichtung gemäß § 71 des Elften Buches Sozialgesetzbuch beschäftigt waren, die Dauer der Maßnahme gegenüber der Regelausbildung um ein Drittel der Ausbildungszeit zu verkürzen.

(4) Auf Antrag soll bei Teilnahme an einer beruflichen Weiterbildung nach dem Dritten Buch Sozialgesetzbuch die Dauer der Maßnahme gegenüber der Regelausbildung verkürzt werden:

1. für Personen gemäß Absatz 1 Nummer 1 um bis zu zwei Drittel der Ausbildungszeit,

2. für Fälle des Absatzes 2 um bis zu zwei Drittel der Ausbildungszeit,
 3. für Personen, die in einem Umfang, der einer Vollzeitbeschäftigung von mindestens zwei Jahren entspricht, in einer Pflegeeinrichtung gemäß § 71 des Elften Buches Sozialgesetzbuch Aufgaben im Bereich der Pflege oder Betreuung wahrgenommen haben, auf der Grundlage einer Kompetenzfeststellung um ein Drittel der Ausbildungszeit.
- (5) Die Verkürzung darf die Durchführung der Ausbildung und die Erreichung des Ausbildungszieles nicht gefährden.
- (6) Die Absätze 1 bis 5 gelten für die Ausbildung nach § 4 Abs. 5 entsprechend.“

2.2 Auswirkungen der Gesetzesänderung

Im Rahmen der stationären und ambulanten Altenpflege stehen die nicht formell erworbenen Kompetenzen von ungelernten Hilfskräften immer mehr im Fokus des Interesses. Denn durch das Gesetz der Bundesregierung zur Stärkung der beruflichen Aus- und Weiterbildung in der Altenpflege vom 13. März 2013 wurde die Möglichkeit zur Anerkennung von informell und non-formal erworbenen Kompetenzen zur Verkürzung der Altenpflegeausbildung gesetzlich festgeschrieben. Aufgrund der Änderung von § 7 AltPflG können Bewerber ohne fachlich einschlägige Vorqualifikationen auf der Grundlage einer Kompetenzfeststellung eine Verkürzung der Ausbildung um ein Drittel ermöglicht bekommen. Das Kompetenzfeststellungsverfahren soll die Anerkennung von informell und non-formal erworbenen Kompetenzen gewährleisten und dabei den gängigen Qualitätsansprüchen eines Messverfahrens genügen (vgl. Blumenauer 2013, S. 1 ff.). Es sollte zudem eine Vielzahl unterschiedlicher Kompetenzen erfassen können; neben den fachlichen Kompetenzen sollten auch fachübergreifende Schlüsselkompetenzen Berücksichtigung finden (vgl. Blumenauer 2013, S. 1 ff.). Zu diesen Schlüsselkompetenzen zählen unter anderem die Basiskompetenzen wie Lesen und Rechnen (vgl. OECD 2005, S. 12) und das mathematische Textverständnis.

2.3 Kompetenzfeststellungsverfahren im Rahmen der Novellierung des Altenpflegegesetzes

Die Entwicklung und Organisation der Kompetenzfeststellungsverfahren zur Verkürzung der Altenpflegeausbildung obliegt nach § 26 Abs. 2 AltPflG

den zuständigen Landesbehörden (vgl. Blumenauer 2013, S. 1). Das Institut für berufliche Bildung, Arbeitsmarkt und Sozialpolitik (INBAS) führte im Rahmen des Programms „Perspektive Berufsabschluss“ des Bundesministeriums für Bildung und Forschung ein Projekt durch, in dem zwei Messverfahren zur Kompetenzbilanzierung erprobt wurden. Das Projekt wurde in Rheinland-Pfalz und Niedersachsen durchgeführt. Das Vorhaben verfolgte das Ziel, durch eine Kompetenzbilanzierung festzustellen, ob die Erreichung des Ausbildungsziels bei Verkürzung der Ausbildungszeit zu erwarten ist. Die Messverfahren haben den Anspruch, die Kompetenzen auf fachlichem und qualitativ hohem Niveau zu bilanzieren. Das Verfahren zur Kompetenzfeststellung (siehe Tabelle 1) basiert auf der Ausbildungs- und Prüfungsverordnung für den Beruf der Altenpflege (vgl. Blumenauer 2013, S. 1 ff.). Im Rahmen des Kompetenzbilanzierungsverfahrens (vgl. Blumenauer 2013, S. 1 ff.) findet die Messung von Schlüsselkompetenzen (vgl. OECD 2005) respektive die Erfassung des mathematischen Textverständnisses, offensichtlich keine Berücksichtigung, obwohl sie aber innerhalb eines solchen Verfahrens erfasst werden sollten. Dies könnte unter anderem daran liegen, dass die Basiskompetenzen als Grundvoraussetzung angenommen werden, sobald ein Bewerber über einen für die Altenpflegeausbildung erforderlichen Schulabschluss verfügt.

In dieser Arbeit wird daher ein Instrument zur Erfassung des mathematischen Textverständnisses (TeMaTex) im Kontext der Altenpflegeausbildung auf seine wissenschaftliche Güte hin überprüft. Dies geschieht mittels geeigneter testtheoretischer Messverfahren. Der Nutzen des Vorhabens liegt darin, dass die entsprechenden Behörden der Länder zukünftig bei der Entscheidungsfindung hinsichtlich einer Verkürzung der Ausbildungszeit auf ein verlässliches Messverfahren zurückgreifen könnten, welches das mathematische Textverständnis erfasst.

Tabelle 1: Verfahren zur Kompetenzbilanzierung in Rheinland-Pfalz (nach Blumenauer 2013, S. 3 f.)

Phase I: Information, Beratung, Prüfung der formalen Voraussetzungen			
Ziele / Inhalt	Beratung und Information von Verantwortlichen in Pflegeeinrichtungen und interessierten Hilfskräften		
	Information zum Bewerbungsverfahren und zur Kompetenzbilanzierung		
	Sichtung der Bewerbungen und Prüfung der formalen Zugangsvoraussetzungen		
Phase II: Kompetenzbilanzierung Teil 1 - schriftlicher und mündlicher Teil			
Setting	In der Altenpflegeschule		
Beteiligte bei der Erprobung	3 Lehrkräfte der Altenpflegeschule 2 Projektmitarbeiterinnen mit pflegerischer Qualifikation 1 Vertreterin der Aufsichts- und Dienstleistungsdirektion Rheinland-Pfalz		
Räumliche Ressourcen	Mindestens 4 Räume (2 Räume für parallele Durchführung von 2 mündlichen Teilen + 1 Raum zur Vorbereitung auf den mündlichen Teil + 1 Aufenthaltsraum für die wartenden Teilnehmerinnen und Teilnehmer)		
Ziele / Inhalt	Methode	Gestaltung / Zielsetzung	Umfang / Dauer
	Fragebogen zur Selbsteinschätzung	Selbstbeurteilung der Teilnehmenden im Hinblick auf die abzubildenden Kompetenzen Einstimmung auf das Verfahren durch Anregung zur Selbstreflexion	30 Fragen zum Ankreuzen Bearbeitungszeit 30 Minuten
	Schriftlicher Teil	Schriftliche Beantwortung von Fragen zu einem entsprechenden Fall mit starkem Praxisbezug. Fall und Fragestellungen wurden auf Grundlage einer Analyse des Kompetenzprofils von un- und angelernten Hilfskräften entwickelt. Jede Frage wurde mit einem Erwartungshorizont hinterlegt. Die Beurteilung erfolgte zu jeder Frage anhand einer entsprechenden Bewertungsmatrix die zwischen 6 Ausprägungsgraden der jeweiligen Kompetenz unterscheiden lässt.	7 Fragen zu unterschiedlichen Themengebieten 15 Minuten Vorbereitungs- und Einlesezeit 60 Minuten Bearbeitungszeit
	Mündlicher Teil	Mündliche Beantwortung von Fragen zu dem gleichen Fall wie im schriftlichen Teil Erwartungshorizont und Beurteilungsmatrix wie oben.	7 Fragen 15 Minuten Vorbereitungszeit 30 Minuten Gesprächszeit

	Auswertung und Entscheidung	Auswertung der Beurteilungen der einzelnen Teile und Entscheidung über Zulassung zum Teil 2 der Kompetenzbilanzierung	Je nach Teilnehmerzahl
Phase III: Kompetenzbilanzierung Teil 2 - praktischer Teil			
Setting	In der Pflegeeinrichtung oder in der Häuslichkeit der zu Pflegenden		
Beteiligte bei der Erprobung	1 Lehrkraft der Altenpflegeschule, 1 Projektmitarbeiterin, 1 Vertreterin der Aufsichts- und Dienstleistungsdirektion (exemplarisch bei einer Teilnehmenden), ggf. Praxisanleiter oder Praxisanleiterinnen aus der Pflegeeinrichtung		
Ziele / Inhalt	Methode	Gestaltung / Zielsetzung	Umfang / Dauer
	Durchführung einer Körperpflege	<p>Im Vorfeld Planung mit der Einrichtung und Einverständnis der zu Pflegenden notwendig</p> <p>Beobachtung des direkten Umgangs und der Kommunikation mit einer zu pflegenden Person</p> <p>Beobachtung der vorhandenen Kompetenzen in der direkten Interaktion mit einer zu pflegenden Person</p>	<p>15 Minuten Vorbereitungsgespräch (Absprachen zur Vorgehensweise, Klärung von offenen Fragen)</p> <p>45 Minuten Durchführung einer Körperpflege</p> <p>15 Minuten Reflexionsgespräch im Anschluss</p>
	Auswertung und Entscheidung	<p>Auswertung und Beurteilung anhand der Beobachtungen und der Bewertungsmatrix</p> <p>Entscheidung über eine Empfehlung zur Verkürzung der Ausbildung</p> <p>Weiterleitung der Empfehlung an die zuständige Stelle zur formalen Prüfung und Information an die Teilnehmenden über weiteres Vorgehen</p>	<p>Ca. 30 Minuten Auswertung</p> <p>Zeiträume bis zur Entscheidung durch die zuständige Stelle sind unterschiedlich</p>

2.4 Zusammenfassende Betrachtung des zweiten Kapitels

In Kapitel 2 wurde erläutert, welche Auswirkungen die Novellierung des Altenpflegegesetzes auf die Altenpflegeausbildung, insbesondere auf die Gewinnung von potentiellen Altenpflegeschülern haben wird. Es konnte aufgezeigt werden, dass anhand der exemplarisch dargestellten Kompetenzfeststellungsverfahren keine ausreichende Möglichkeit besteht, für die Ausbildung relevante Basiskompetenzen respektive Kulturtechniken wie das mathematische Textverständnis erfassen zu können.

3 Mathematisches Berufsprofil der Altenpflege

Altenpfleger arbeiten in Alten- und Pflegeheimen, bei ambulanten Altenpflege- und Altenbetreuungsdiensten, in geriatrischen und gerontopsychiatrischen Abteilungen von Krankenhäusern, in Hospizen und in Pflege- und Rehabilitationskliniken. Zu ihren Hauptaufgaben gehören die Pflege, die Betreuung und die Beratung hilfsbedürftiger älterer Menschen und deren Angehöriger. Altenpfleger unterstützen hilfsbedürftige ältere Menschen bei den Verrichtungen des täglichen Lebens, etwa bei der Körperpflege, beim Essen oder beim Anziehen. Die Führung von Gesprächen über persönliche Angelegenheiten, das Motivieren zu einer aktiven Freizeitgestaltung und die Begleitung bei Behördengängen oder Arztbesuchen gehören zu den Aufgabengebieten eines Altenpflegers. Im Rahmen der ambulanten Pflege arbeiten Altenpfleger eng mit Angehörigen zusammen und schulen diese in Pflege- und Technik. Dazu zählen auch therapeutische und medizinisch-pflegerische Aufgaben, wie Verbände wechseln, Spülungen durchführen und Medikamente verabreichen nach ärztlicher Verordnung. Das direkte Arbeitsumfeld eines Altenpflegers beschränkt sich in erster Linie auf Patientenzimmer, Behandlungsräume, Sanitärräume, Gruppen- und Aufenthaltsräume, Besprechungs- und Büroräume sowie auf Patientenwohnungen. Das vielfältige Aufgabengebiet zeigt, dass in diesem Beruf ein hohes Sorgfalts- und Verantwortungsbewusstsein, wie etwa beim exakten Einhalten von Pflegemaßnahmen oder bei deren Dokumentation, vorausgesetzt werden muss. Darüber hinaus ist eine gute körperliche Konstitution wie beim Umbetten von Patienten empfehlenswert. Ein hohes Einfühlungsvermögen im Umgang mit Patienten und Angehörigen sowie eine hohe psychische Stabilität, welche im Umgang mit schwer kranken oder sterbenden Patienten erforderlich ist, sind notwendig. Des Weiteren wird ein Interesse an Sozialkunde, Religionslehre/Ethik, Deutsch, Wirtschaft und Mathematik verlangt. Für die Ausbildung wird gewöhnlich ein mittlerer Bildungsabschluss vorausgesetzt. Die Berufsfachschulen wählen Bewerber nach eigenen Kriterien aus. In Einrichtungen, die ihre Entgelte nach den Tarifbedingungen des öffentlichen Dienstes gestalten, erhalten die Auszubildenden beispielsweise folgende

Entgelte (monatlich brutto): 1. Ausbildungsjahr: € 976; 2. Ausbildungsjahr: € 1.037; 3. Ausbildungsjahr: € 1.138 (vgl. Bundesagentur für Arbeit 2016).

In Kapitel 3.1 erfolgt eine Analyse mathematischer Anforderungen im curricularen Bereich der Altenpflegeausbildung. Kapitel 3.2 beschäftigt sich mit der Analyse mathematischer Anforderungen anhand ausgewählter Aufgaben aus der theoretischen Ausbildung. In Kapitel 3.3 werden die zentralen mathematischen Anforderungen für den Altenpflegeberuf dargestellt. Abschließend erfolgt in Kapitel 3.4 eine Zusammenfassung.

3.1 Analyse mathematischer Anforderungen im curricularen Bereich

Der Unterricht in der Berufsfachschule basiert auf den Rahmenlehrplänen der jeweiligen Bundesländer. Die jeweiligen Rahmenlehrpläne orientierten sich an den geltenden gesetzlichen Regelungen des Bundes – dem bundeseinheitlichen Altenpflegegesetz und der dazugehörigen Ausbildungs- und Prüfungsverordnung aus dem Jahre 2003. Die Lehrpläne sind nach dem Lernfeldkonzept konzipiert. Exemplarisch soll im Folgenden der Rahmenlehrplan Hessen näher betrachtet werden. Laut Ausbildungs- und Prüfungsverordnung ist die Altenpflegeausbildung in vier große Lernbereiche mit insgesamt vierzehn Lernfeldern eingeteilt.

Tabelle 2: Lernbereiche (in Anlehnung an das Hessische Sozialministerium 2011, S. 11)

Lernbereiche	Titel	Stundenumfang
Lernbereich 1	Aufgaben und Konzepte in der Altenpflege	1200 Stunden
Lernbereich 2	Unterstützung alter Menschen bei der Lebensgestaltung	300 Stunden
Lernbereich 3	Rechtliche und institutionelle Rahmenbedingungen altenpflegerischer Arbeit	160 Stunden
Lernbereich 4	Altenpflege als Beruf	240 Stunden
		+ 200 Stunden zur freien Verfügung
	Summe:	2100 Stunden

Tabelle 2 beziehungsweise Tabelle 3 zeigt eine Übersicht zu den einzelnen Lernbereichen respektive Lernfeldern inklusive der in der dreijährigen Ausbildungszeit zu erbringenden Unterrichtsstunden am Lernort Schule.

Tabelle 3: Lernfelder (in Anlehnung an das Hessische Sozialministerium 2011, S. 12 ff.)

Lernbereich 1: Aufgaben und Konzepte in der Altenpflege		
Lernfelder	Titel	Stundenumfang
Lernfeld 1.1	Theoretische Grundlagen in das Altenpflegerische Handeln einbeziehen	80 Stunden
Lernfeld 1.2	Pflege alter Menschen planen, durchführen, dokumentieren und evaluieren	120 Stunden
Lernfeld 1.3	Alte Menschen personen- und situationsbezogen pflegen	720 Stunden
Lernfeld 1.4	Anleiten, beraten und Gespräche führen	80 Stunden
Lernfeld 1.5	Bei der medizinischen Diagnostik und Therapie mitwirken	200 Stunden
	Zwischensumme:	1200 Stunden
Lernbereich 2: Unterstützung alter Menschen bei der Lebensgestaltung		
Lernfelder	Titel	Stunden
Lernfeld 2.1	Lebenswelten und soziale Netzwerke alter Menschen beim Altenpflegerischen Handeln berücksichtigen	120 Stunden
Lernfeld 2.2	Alte Menschen bei der Wohnraum- und Wohnumfeldgestaltung unterstützen	60 Stunden
Lernfeld 2.3	Alte Menschen bei der Tagesgestaltung und bei selbst organisierten Aktivitäten unterstützen	120 Stunden
	Stunden:	300 Stunden
Lernbereich 3: Rechtliche und institutionelle Rahmenbedingungen Altenpflegerischer Arbeit		
Lernfelder	Titel	Stunden
Lernfeld 3.1	Institutionelle und rechtliche Rahmenbedingungen beim Altenpflegerischen Handeln berücksichtigen	120 Stunden
Lernfeld 3.2	An qualitätssichernden Maßnahmen in der Altenpflege mitwirken	40 Stunden
	Stunden:	160 Stunden

Lernbereich 4: Altenpflege als Beruf		
Lernfelder	Titel	Stunden
Lernfeld 4.1	Berufliches Selbstverständnis entwickeln	60 Stunden
Lernfeld 4.2	Lernen lernen	40 Stunden
Lernfeld 4.3	Mit Krisen und schwierigen sozialen Situationen umgehen	80 Stunden
Lernfeld 4.4	Die eigene Gesundheit erhalten und fördern	60 Stunden
	Summen:	240 Stunden

Tabelle 4 zeigt die Lernfelder auf, in denen mathematische Grundqualifikationen gefordert werden.

Tabelle 4: Relevante Lernfelder

Lernbereiche	Nummern der Lernfelder
Lernbereich 1	1.1, 1.3, 1.5
Lernbereich 2	2.1
Lernbereich 3	3.1
Lernbereich 4	---

Im Folgenden sollen die mathematischen Grundqualifikationen der dargestellten Lernbereiche und Lernfelder näher beschrieben werden. Im Lernbereich 1 *Aufgaben und Konzepte in der Altenpflege* mit den Lernfeldern 1.1, 1.3 und 1.5 sind mathematische Inhalte integriert. Ein Teilziel im Lernfeld 1.1 ist die „Mitarbeit bei der Umsetzung von Ergebnissen der Pflegeforschung in der Einrichtung“ (Hessisches Sozialministerium 2011, S. 14). Hier wird von den Auszubildenden gefordert, dass sie wissenschaftliche Texte lesen und verstehen können. Damit die Auszubildenden dieses Ziel erreichen, ist es wichtig, dass sie die Grundlagen über statistische Auswertungsverfahren kennen. Dazu zählen unter anderem die Berechnungen statistischer Maßzahlen zur Kennzeichnung einer Häufigkeitsverteilung (vgl. Gierse 2001, S. 9).

Ein Teilziel des Lernfeldes 1.3 *Alte Menschen personen- und situationsbezogen pflegen* ist unter anderem, Indikationen „unterschiedlicher Darreichungsformen (z.B. s.c. Infusionen, Tabletten, Aerosole, Suppositorien, Sal-

ben, Tropfen, Pflaster etc.), Dosierungen, sowie Wirkweisen und Wechselwirkungen“ (Hessisches Sozialministerium 2011, S. 19) zu erkennen. Dies umfasst unter anderem die Berechnung von Medikamenten-Dosierungen sowie die Berechnung der Einlaufgeschwindigkeit und Einlaufmenge von Infusionslösungen. Ein weiteres Teilziel in diesem Lernfeld stellt der „Umgang mit Hygiene und Desinfektion unter betriebswirtschaftlichen und ökologischen Aspekten“ dar (Hessisches Sozialministerium 2011, S. 20). Hierbei wird hervorgehoben, dass die Auszubildenden Kenntnisse besitzen, um entsprechende „Berechnungen im Zusammenhang mit Desinfektionslösungen“ (Gierse 2001, S. 8) anstellen zu können. Darüber hinaus sollen die Auszubildenden „die Bestandteile von Nahrungsmitteln und deren Verarbeitung zu einer ausgewogenen Ernährung“ (Hessisches Sozialministerium 2011, S. 21) kennen lernen sowie „verschiedene Zusammensetzungen von Diäten für verschiedene Krankheitsbilder“ (Hessisches Sozialministerium 2011, S. 21) herausarbeiten können. Hierzu zählen Berechnungen in Bezug auf die Ernährung bei Diabetes mellitus. Des Weiteren ist ein Teilziel im Lernfeld 1.3, dass die Auszubildenden die Vitalfunktionen wie Blutdruck, Puls, Körpertemperatur und Atmung messen und im Anschluss die Werte beurteilen können (vgl. Hessisches Sozialministerium 2011, S. 30).

Das Lernfeld 1.5 *Bei der medizinischen Diagnostik und Therapie mitwirken* hat unter anderem zum Ziel, „Maßnahmen der Behandlungspflege und ihre Anwendung in der Praxis“ (Hessisches Sozialministerium 2011, S. 46) zu erlernen. Hier sind zum Beispiel Berechnungen bezüglich einer Insulintherapie bei einem Patienten mit Diabetes mellitus von Interesse.

Im Lernfeld 2.1 *Lebenswelten und soziale Netzwerke alter Menschen beim pflegerischen Handeln berücksichtigen* soll folgendes Teilziel erreicht werden: „Auswirkungen der demographischen Entwicklung auf die eigene Ausbildungseinrichtung einschätzen können (Hessisches Sozialministerium 2011, S. 50). Im Rahmen dieses Lernfeldes wird von den Auszubildenden die Fähigkeit in Bezug auf das Lesen und die Interpretation von Diagrammen zu Altersschichtungen gefordert.

Im Lernfeld 3.1 *Institutionelle und rechtliche Rahmenbedingungen beim Altenpflegerischen Handeln berücksichtigen* sollen die Auszubildenden folgende Teilziele erreichen: „Kennen lernen der Zusammenhänge von Personalkosten, Sachkosten und wirtschaftlichem Pflegehandeln ... Kenntnisse zur Finanzierung von Leistungen im stationären und ambulanten Bereich erlangen ... Kennen lernen der Abrechnungssysteme in der Pflege (insbesondere beim Ausbildungsträger) ... Mitwirkung bei der Vorbereitung und dem Verfahren der Eingruppierung durch den MDK ... Lernen Pflegezeit individuell zu planen und zu kontrollieren“ (Hessisches Sozialministerium 2011, S. 63). Hierbei sind unter anderem Kenntnisse auf dem Gebiet der Lohn- und Gehaltsabrechnung notwendig. In dem Lernbereich 4 *Altenpflege als Beruf* mit den Lernfeldern 4.1, 4.2, 4.3 und 4.4 konnte aus dem Rahmenlehrplan kein mathematischer Inhalt analysiert werden.

Zusammenfassend lässt sich feststellen, dass mathematische Grundqualifikationen in vielen Lernfeldern gefordert werden, insbesondere in den Lernfeldern 1.1 *Theoretische Grundlagen in das Altenpflegerische Handeln einbeziehen*, 1.3 *Alte Menschen personen- und situationsbezogen pflegen*, 1.5 *Bei der medizinischen Diagnostik und Therapie mitwirken*, 2.1 *Lebenswelten und soziale Netzwerke alter Menschen beim pflegerischen Handeln berücksichtigen* und 3.1 *Institutionelle und rechtliche Rahmenbedingungen beim Altenpflegerischen Handeln berücksichtigen*.

3.2 Analyse mathematischer Anforderungen anhand ausgewählter Aufgaben aus der theoretischen Ausbildung

Nach der Einführung des Altenpflegegesetzes und der Ausbildungs- und Prüfungsverordnung sowie der Umstellung der Rahmenlehrpläne auf das Lernfeldkonzept sind die mathematischen Anforderungen für den Altenpflegeberuf nicht auf den ersten Blick aus den Dokumenten ersichtlich. Die mathematischen Grundqualifikationen sind in die einzelnen Lernbereiche beziehungsweise Lernfelder integriert (siehe Kapitel 3.1). Um nun einen Überblick zu erhalten, welche mathematischen Grundqualifikationen in der Altenpflegeausbildung konkret gefordert werden, wird im weiteren Verlauf dieser Arbeit eine Analyse anhand ausgewählter Aufgaben aus einem Buch

zur Prüfungsvorbereitung durchgeführt. Im Folgenden sollen die Schritte der Analyse exemplarisch an einer Aufgabe aufgezeigt werden.

Übersicht zur Fallsituation: Bei dieser Fallsituation wird dem Auszubildenden ein Fließtext mit dazugehöriger Tabelle vorgelegt. Im Text wird eine alkoholranke Frau beschrieben, die an einer Bauchspeicheldrüsenentzündung (Pankreatitis) leidet und einen stark reduzierten Allgemein- und Ernährungszustand aufweist. Zu diesem Text müssen drei Aufgaben von den Auszubildenden bearbeitet werden. In Tabelle 5 werden die drei Aufgaben nach den jeweils anzuwendenden Rechenoperationen dargestellt.

Tabelle 5: Analyse der Aufgaben 1 - 3 zur Fallsituation: Parenterale Ernährung

	Mathematische Anforderungen	Rechenoperationen
Aufgabe 1	Arithmetik: Addition	1500 ml + 500 ml
Aufgabe 2	Prozentrechnung: Prozentwert berechnen	$P = \frac{1477 \text{ ml} * 3,4}{100}$
Aufgabe 3	I. Prozentrechnung: Prozentwerte berechnen	$P = \frac{1477 \text{ ml} * 2,0}{100}$ $P = \frac{1477 \text{ ml} * 2,7}{100}$
	II. Arithmetik: Multiplikation	29,54 ml * 24 h
	III. Arithmetik: Addition	708,96 ml + 39,88 ml

Fallsituation zur parenteralen Ernährung: Frau Möhlich ist 49 Jahre alt. Seit ihrer Heirat erledigte sie ausschließlich den gemeinsamen Haushalt. Sie begann nach und nach Alkohol zu trinken und war immer öfter betrunken. Vor 14 Jahren reichte ihr Mann die Scheidung ein, da er sich in der Ehe überfordert fühlte. Frau Möhlich schaffte es nicht mehr, die Wohnung in Ordnung zu halten. Sie zahlte Rechnungen und die Miete nicht mehr termingerecht, sodass es zur Kündigung kam. Seitdem lebt Frau Möhlich auf der Straße, hin und wieder bittet sie in einer Obdachlosenunterkunft um kurzfristige Aufnahme. Vorletzte Nacht wachte Frau Möhlich dort mit sehr starken Bauchschmerzen auf, die bis zum Morgen nicht nachließen. Im Krankenhaus erzählte sie, dass diese starken Schmerzen schon länger bestehen und dass sie oft Durchfall habe.

Es wird eine Bauchspeicheldrüsenentzündung (Pankreatitis) diagnostiziert. Der Allgemein- und der Ernährungszustand sind deutlich reduziert. Nach einem Tag auf der Intensivstation wird Frau Möhlich auf die internistische Station in ein Drei-Bett-Zimmer verlegt. Sie wirkt geschwächt und ängstlich. Die Haut ist trocken, teilweise schuppig und ungepflegt. Das Gewicht von Frau Möhlich beträgt 50 kg. Bei ihr wird eine Einfuhrkontrolle angelegt. Frau Möhlich erhält täglich eine Kanne Tee (1000 ml) und zusätzlich eine Flasche Wasser (500 ml). Darüber hinaus erhält sie über 24 Stunden kontinuierlich eine parenterale Ernährung mit SmofKabiven® (1477 ml). Hierfür wurde ein Zentraler Venenkatheter (ZVK) gelegt. Nach zwei Wochen wiegt Frau Möhlich 60 kg.

Bearbeiten Sie die nachfolgenden Aufgaben zu diesem Text. Beachten Sie dazu die folgende Tabelle.

Tabelle 6: Infusionsgeschwindigkeit in ml/h

SmofKabiven® zentral bei kontinuierlicher Zufuhr über 24 h				
Gewicht	50 kg	60 kg	70 kg	80 kg
Tag 1	1,7 %	2,0 %	2,4 %	2,7 %
Tag 2	2,4 %	2,7 %	3,0 %	3,4 %
Tag 3	2,7 %	3,4 %	3,7 %	4,1 %
Tag 4	3,4 %	4,1 %	4,4 %	4,7 %

Aufgabe 1: Berechnen Sie die tägliche Gesamttrinkmenge von Frau Möhlich, wenn sie den Inhalt der Kanne Tee und der Flasche Wasser vollständig zu sich nehmen würde.

Analyse zur Aufgabe 1: In der ersten Aufgabe muss eine einfache mathematische Rechenoperation (Addition) durchgeführt werden. Die Auszubildenden sollen berechnen, wie die Gesamttrinkmenge der Patientin lautet, wenn sie den Inhalt der Kanne Tee und der Flasche Wasser vollständig zu sich nimmt. Um diese Aufgabe lösen zu können, müssen im Fließtext die zwei Mengenangaben 1000 ml und 500 ml gefunden und diese zusammen-

gezählt werden. Die Mengenangabe 1000 ml wird im Text und in der Aufgabe mit dem Begriff „Kanne Tee“ und die Mengenangabe 500 ml wird mit dem Begriff „Flasche Wasser“ codiert.

Aufgabe 2: Frau Möhlich erhält im Rahmen der parenteralen Ernährung kontinuierlich über 24 h SmofKabiven® (1477 ml). Berechnen Sie die Menge an SmofKabiven®, die Frau Möhlich am vierten Tag stündlich erhält.

Analyse zur Aufgabe 2: Die zweite Aufgabe ist aus dem Bereich des Prozentrechnens. Basierend auf einem Grundwert von 1477 ml soll der entsprechende Prozentwert in Form einer Milliliterangabe eines Produktes zur parenteralen Ernährung berechnet werden. Der Auszubildende erhält die Aufgabe, aus einer Tabelle den richtigen Prozentsatz herauszulesen und darauf aufbauend eine zusammenhängende arithmetische Rechenoperation (Multiplikation und Division) durchzuführen.

Aufgabe 3: Frau Möhlich wiegt nach zwei Wochen 60 kg. Sie erhält am zweiten Tag stündlich mehr SmofKabiven®. Insgesamt erhielt Frau Möhlich 748,84 ml SmofKabiven®. Berechnen Sie die Menge an SmofKabiven®, welche Frau Möhlich nach dem ersten Tag erhielt.

Analyse zur Aufgabe 3: Bei Aufgabe 3 handelt es sich um eine Aufgabe aus den Bereichen des Prozentrechnens und der Arithmetik. Hierbei muss die Gesamtmenge an SmofKabiven® ermittelt werden, welche Frau Möhlich nach 24 Stunden erhalten hat. Zuerst muss es dem Auszubildenden gelingen, aus der Tabelle die entsprechenden Prozentsätze herauszulesen, um dann die Prozentwerte zu bestimmen. Danach muss der Prozentwert des ersten Tages mit 24 multipliziert und mit dem Prozentwert des zweiten Tages addiert werden.

3.3 Mathematische Anforderungsbereiche in der Ausbildung

Für die Erstellung des Berufprofils wurde ausschließlich auf das Buch zur Prüfungsvorbereitung *Prüfungswissen Altenpflege* aus dem Elsevier-Verlag zurückgegriffen, da das dazugehörige Lehrbuch *Altenpflege heute* in vielen

Altenpflegeschulen als Grundlagen-Lehrbuch für die theoretische Ausbildung verwendet wird. Die Inhalte des Lehrbuchs und des Buchs zur Prüfungsvorbereitung haben den Vorteil, dass sie die Ausbildungsinhalte nach den dazugehörigen Lernbereichen und Lernfeldern strukturieren. Basierend auf der Analyse ausgewählter Aufgaben aus dem zugrundeliegenden Buch konnte für die Altenpflegeausbildung ein mathematisches Anforderungsprofil erstellt werden. Die Analyse umfasste 1416 Aufgaben, von denen lediglich 9 Aufgaben mathematische Anforderungen beinhalten. Dies macht einen mathematischen Anteil in der theoretischen Altenpflegeausbildung von etwa 0,6 % aus. Jede der 9 Aufgaben wurde nach der in Kapitel 3.2 erläuterten Vorgehensweise analysiert und anschließend in Kategorien eingeteilt, welche auch in dem Projekt Mathe-Meister (vgl. Stein & Winter 2009) Anwendung fanden. Hierzu zählen die Kategorien Arithmetik, Runden, Einheiten, Algebra, Geometrie, Diagramme und Tabellen sowie Dreisatz und Prozentrechnung (vgl. Stiftung Rechnen 2015, S. 60). Eine Aufgabe wurde in zwei Kategorien gleichzeitig eingeordnet. Es fanden innerhalb der 9 Aufgaben 22 Zuordnungen zu den Kategorien beziehungsweise mathematischen Fähigkeiten statt. Insgesamt sind für die Ausbildung zum Altenpfleger die Kategorien *Arithmetik* sowie *Tabellen und Diagramme* von großer Bedeutung. Die Kategorie *Arithmetik* stellt mit etwa 67,7 % das größte mathematische Themengebiet in der Altenpflegeausbildung dar. Es folgt die Kategorie *Tabellen und Diagramme* mit einem Aufgabenanteil von 33,3 %.

Im Folgenden werden die beiden Kategorien näher betrachtet. Es wird dargestellt, welche spezifischen mathematischen Teilgebiete durch die einzelnen Aufgaben in den jeweiligen Kategorien abgebildet werden. Hierfür werden die jeweiligen Aufgaben eines mathematischen Teilgebietes gezählt.

In der Kategorie *Arithmetik* stellt die Addition mit 8 von 18 Zuordnungen den größten Anteil in dieser Kategorie dar. Dem folgt die Multiplikation mit 7 von 18 Zuordnungen an. Die Division ist mit 3 von 18 Zuordnungen vertreten. Insgesamt werden alle der Kategorie *Arithmetik* zugeordneten Aufgaben im mathematischen Teilgebiet *Grundrechenarten* verortet. Die anderen Teilgebiete wie *Rechengesetze*, *Rechnen mit negativen Zahlen* und *Rechnen mit*

Potenzen (vgl. Stiftung Rechnen 2015, S. 61) sind in keiner der Aufgaben von Bedeutung.

In der Kategorie *Tabellen und Diagramme* nehmen die Tabellen den größeren Schwerpunkt mit 2 von 3 Zuordnungen ein. Abbildung 1 stellt beispielhaft eine Aufgabe aus dem Prüfungsbuch dar, welche sowohl in die Kategorie *Arithmetik* als auch in die Kategorie *Tabellen und Diagramme* einzuordnen war.

Fallbeispiel

Frau Rosa Weber, 87 Jahre alt, lebt seit 3 Jahren im Altenheim am Wiesengrund. Bis zu ihrem Heimeinzug lebte sie zu Hause bei ihrer Tochter. Sie wurde daheim zunehmend vergesslicher, aber auch aggressiver, und da sie keine körperlichen Einschränkungen hatte, brachte sie sich oft in gefährliche Situationen. So kochte sie z. B. und vergaß, die Herdplatte auszuschalten, oder sie ging einkaufen und fand nicht mehr nach Hause. Der Hausarzt diagnostizierte eine Alzheimererkrankung. Da die Tochter noch berufstätig ist, war sie bald mit der Situation überfordert und brachte ihre Mutter ins Altenheim.

Frau Weber isst gerne und – obwohl sie Diabetikerin ist – besonders gerne Süßigkeiten. Ihr fehlt die Einsicht in eine notwendige Diät. An den Füßen sind schon leichte Sensibilitätsstörungen aufgetreten; eine kleine Verletzung durch die Fußpflegerin heilte schlecht ab. Bei einer Größe von 1,70 m wiegt sie 82 kg. Sie trinkt relativ wenig. Bis vor 5 Jahren hat sie geraucht. Sie leidet häufiger an Infektionen der oberen Atemwege. Im Moment hat sie keine Beschwerden.

Bisher war Frau Weber den ganzen Tag in Bewegung. Vor 5 Tagen jedoch stürzte sie über einen am Boden liegenden Schuh und prellte sich die rechte Schuler. Seitdem hat sie Schmerzen bei jeder Bewegung im Brust- und Schulterbereich und weigert sich, aus dem Bett aufzustehen. Sie liegt meist im Bett und steht nur manchmal auf, wenn sie auf die Toilette muss, manchmal nässt sie aber auch ein.

Abbildung 1: Aufgabe zur Prüfungsvorbereitung – Fallbeispiel (nach Derrer-Merk 2013, S. 156 f.)

Aufgabe:

Durch die veränderte Pflegesituation müssen Sie die Pflegeplanung von Frau Weber neu erstellen. Für die aktuelle Pflegeplanung überprüfen Sie anhand der modifizierten Atemskala nach C. Bienstein die momentane Pneumoniegefährdung von Frau Weber. Auf welche Punktzahl kommen Sie? Bewerten Sie das Ergebnis!

Abbildung 2: Aufgabe zur Prüfungsvorbereitung - Aufgabenstellung (nach Derrer-Merk 2013, S. 161)

Tabelle 7: Aufgabe zur Prüfungsvorbereitung - Tabelle zur Aufgabenstellung (nach Derrer-Merk 2013, S. 161 f.)

Kriterium	Punkte	Bewertung	
Bereitschaft zur Mitarbeit		0 Kontinuierliche Mitarbeit 1 Mitarbeit nach Aufforderung	2 Nur nach Aufforderung 3 Keine
Vorliegende Atemwegserkrankungen		0 Keine 1 Leichter Infekt im Nasen-/Rachenraum	2 Bronchialinfekt 3 Lungenerkrankung
Frühere Lungenerkrankungen		0 Keine 1 Leichte, z. B. bronchopulmonale grippale Infekte 2 Schwere Verläufe 3 Schwere Lungenerkr. mit bleibender Atemfunktions-einschränkung	
Immunschwäche		0 Keine 1 Leicht (z.B. lokale Infektion)	2 Erhöht 3 Völlig
Raucher/Passivraucher		0 Nichtraucher, geringfügiges Passivrauchen 1 Pro Tag 6 Zigaretten mit niedrigem Teer-/Kondensatgehalt < 10 mg oder regelmäßiges Passivrauchen 2 Pro Tag 6 Zigaretten mit 10 – 13 mg Teer-/Kondensatgehalt oder regelmäßiges Passivrauchen (z. B. bei Rauchen des Partners) 3 Intensives Rauchen, mehr als 6 Zigaretten mit > 15 mg Teer-/Kondensatgehalt, ständiger passiver Raucherkonsum	
Schmerzen		0 Keine 1 Leichte Schmerzen, Dauerschmerzen 2 Mäßige atmungsbeeinflussende Schmerzen 3 Starke atmungsbeeinflussende Schmerzen	
Schluckstörungen		0 Keine 1 Bei flüssiger Nahrung	2 Bei breiiger Nahrung 3 Komplette Schluckstörungen, auch beim Schlucken von Speichel

Manipulative orotracheale Maßnahmen		0 Keine 1 Pflegemaßnahmen, z.B. Nasen- und Mundpflege 2 Zusätzlich orale oder nasale Absaugung 3 Zusätzlich endotracheale Absaugung ohne oder mit liegendem Tubus	
Mobilitätseinschränkung		0 Keine 1 Eingeschränkte Mobilität, durch Gehhilfen kompensierbar	2 Hauptsächlich Bettruhe 3 Völlige Einschränkung
Arbeit in lungengefährdendem Beruf		0 Keine 1 Für 1 - 2 Jahre	2 Für 2 - 10 Jahre 3 > 10 Jahre
Intubationsnarkose, Beatmung		0 In den letzten drei Wochen keine 1 Kurze Intubationsnarkose (bis 2 Stunden) 2 Langdauernde Intubationsnarkose (> 2 Stunden) 3 Mehrere Intubationsnarkosen oder > 12 Stunden Beatmung	
Bewusstseinslage		0 Keine Einschränkung 1 Leichte Einschränkung (reagiert auf Ansprache folgerichtig) 2 Reagiert auf Ansprache nicht folgerichtig 3 Keine Reaktion	
Atemanstrengung		0 Zwerchfell- und Thoraxatmung ohne Anstrengung 1 Zwerchfell- oder Thoraxatmung mit Anstrengung 2 Zwerchfell- oder Thoraxatmung mit großer Hilfestellung 3 Keine Zwerchfell- oder Thoraxatmung möglich	
Atemfrequenz		0 14 - 20 Atemzüge/Min. 1 Unregelmäßige Atmung 2 Regelmäßige bradypnoische oder tachypnoische Atmung 3 Regelmäßige, sehr tiefe oder auch oberflächliche Atemzüge oder zwischen tachypnoisch und bradypnoisch wechselnde Atmung	
Atemdepressive Arzneimittel		0 Keine 1 Unregelmäßige Einnahme, geringe Atemdepression 2 Regelmäßige Einnahme, mäßige Atemdepression 3 Regelmäßige Einnahme spezifisch atemdepressiver Arzneimittel (z.B. Opiate, Barbiturate)	
Summe		Bewertung: 0 - 6 Punkte = Nicht gefährdet 7 - 15 Punkte = Gefährdet 16 - 45 Punkte = Hochgradig gefährdet, manifeste Atemstörung	

Bei der Aufgabe muss der Auszubildende unter Anwendung der leicht modifizierten Atemskala nach C. Bienstein das Pneumonierisiko von Frau Weber bestimmen. Hierfür muss er relevante Informationen aus dem Fallbeispiel entnehmen und diese Informationen anhand der 15 Kriterien aus der Skala bewerten. Im Folgenden wird die Lösung der Aufgabe dargestellt.

Tabelle 8: Lösung zur Aufgabenstellung (nach Derrer-Merk 2013, S. 163)

Lösung:	
Auswertung	Punkte
Bereitschaft zur Mitarbeit	1
Vorliegende Atemwegserkrankungen	0
Frühere Lungenerkrankungen	1
Immunschwäche	1
Raucher/Passivraucher	0
Schmerzen	1
Schluckstörungen	0
Manipulative orotracheale Maßnahmen	0
Mobilitätseinschränkung	1,5
Arbeit in lungengefährdendem Beruf	0
Intubationsnarkose, Beatmung	0
Bewusstseinslage	0
Atemanstrengung	0
Atemfrequenz	0
Atemdepressive Arzneimittel	1
Summe	6,5
Punktzahl 6,5 → eine Gefährdung liegt vor.	

3.4 Zusammenfassende Betrachtung des dritten Kapitels

In Kapitel 3 wurde erörtert, welche Relevanz das mathematische Textverständnis für die Altenpflegeausbildung hat. Mithilfe der Analysen im Bereich exchange value (vgl. Coben 2002) konnte konstatiert werden, dass das mathematische Textverständnis in unterschiedlichen Lernfeldern des Lehrplanes für die Altenpflegeausbildung an privaten Altenpflegesschulen in Rheinland-Pfalz sowie in der Literatur zur Prüfungsvorbereitung offenbar eine nicht geringfügige Relevanz für die Altenpflegeausbildung darstellt. Unter exchange value werden die (mathematischen) Anforderungen innerhalb der schulischen Altenpflegeausbildung verstanden (vgl. Stein & Winter 2009, S.

2). Im Rahmen der Analyse zum Bereich des exchange value wurden 1416 Aufgaben aus einem Lehrbuch zur Prüfungsvorbereitung für die schulische Altenpflegeausbildung betrachtet. Daher ist festzuhalten, dass die durch die Analyse ermittelten mathematischen Anforderungen durch die Schulcurricula der Sekundarstufe I abgedeckt werden (vgl. Stiftung Rechnung 2015, S. 7). Dies bedeutet, dass Altenpflegeschüler, die die mathematischen Themenbereiche der Sekundarstufe I beherrschen, es wahrscheinlich einfacher haben, den mathematischen Anforderungen in der Altenpflegeausbildung gerecht zu werden.

4 Mathematisches Textverständnis

In Kapitel 4.1 wird beschrieben, was unter dem Konstrukt des mathematischen Textverständnisses zu verstehen ist. Kapitel 4.2 beschäftigt sich mit dem Aufgabentyp zur Messung des mathematischen Textverständnisses. Hierbei werden verschiedene Aufgabentypen und Aufgabenformen dargestellt. In Kapitel 4.3 werden die Bearbeitungs- und Lösungsprozesse von Textaufgaben erläutert. Abschließend erfolgt in Kapitel 4.4 eine Zusammenfassung.

4.1 Begriffsbestimmung

Unter mathematischem Textverständnis werden jene Fähigkeiten zusammengefasst, die benötigt werden, um aus einem mathematikhaltigen Text relevante Informationen von irrelevanten Informationen zu trennen, damit eine sich auf diesen Text beziehende mathematische Textaufgabe gelöst werden kann (vgl. Jordan 2011a, S. 57). Im Rahmen des *International Adult Literacy Survey* (vgl. IALS 2000) wird dies als „Umgang mit Zahlen (‘quantitative literacy’) – das Wissen und die Fertigkeiten, die gebraucht werden, um eine arithmetische Operation, entweder einzeln oder in einer Abfolge, auf in gedrucktem Material enthaltenen Zahlen anzuwenden, wie z. B. die Ermittlung eines Kontostandes oder eines angemessenen Trinkgeldes, das Ausfüllen eines Bestellscheines oder die Berechnung der Höhe von Zinsen auf einen Kredit aus den Angaben einer Anzeige“ (Jordan 2011a, S. 86) verstanden.

Die *IALS-Studie* ist eine international angelegte Untersuchung, die von unterschiedlichen Ländern und Forschungsorganisationen entwickelt und durchgeführt wurde (vgl. OECD 2000, S. 9). An der Entwicklung und Durchführung der Studie waren unter anderem folgende Einrichtungen beteiligt: *Statistics Canada*, *Educational Testing Service* (ETS) und die *Organisation for Economic Cooperation and Development* (OECD). Im Jahre 1994 nahmen neun Länder an dieser weltweit ersten Vergleichsstudie zur Erfassung der Literalität Erwachsener teil, dazu zählten die englischen und französischen Regionen Kanadas, Frankreich, Deutschland, Irland, Niederlande,

Polen, Schweden, die deutsch-französische Schweiz und die USA (vgl. OECD 2000, S. 9). Die Ergebnisse dieser ersten Erhebung wurden 1995 in dem Bericht *Literacy, Economy and Society: Results of the First International Adult Literacy Survey* veröffentlicht. Im Jahre 1997 wurden im Bericht *Literacy Skills for the Knowledge Society: Further Results from the International Adult Literacy Survey* weitere Ergebnisse zur zweiten Erhebungsphase aus fünf Ländern und Kontinenten – dazu zählten Australien, die flämische Gemeinschaft in Belgien, Großbritannien, Neuseeland und Nordirland – veröffentlicht (vgl. OECD 2000, S. 9). In einer dritten Erhebungsphase wurden im Jahre 1998 weitere neun Länder und Regionen untersucht. Zu den untersuchten Ländern und Regionen zählten Chile, die Tschechische Republik, Dänemark, Finnland, Ungarn, Italien, Norwegen, Slowenien und die italienische Schweiz. Die Ergebnisse der dritten Erhebungsphase wurden in dem Bericht *Literacy in the Information Age – Final Report of the international Adult Literacy survey* veröffentlicht (vgl. OECD 2000, S. 9).

Definition Literalität: Im Rahmen der IALS-Studie werden Fähigkeiten unter dem Begriff *Kompetenz* zusammengefasst. Diese Fähigkeiten befinden sich auf einem Fähigkeits-Kontinuum (vgl. OECD 2000, S. 10). Hierbei wird unterschieden, ob ein erwachsener Mensch die Fähigkeit besitzt, lesen und schreiben zu können oder nicht (vgl. OECD 2000, S. 10). Basierend auf diesem Fähigkeits-Kontinuum ist es möglich, aufzuzeigen, auf welchem Niveau Erwachsene Informationen verarbeiten, um in der Gesellschaft und im Arbeitsleben fungieren zu können (vgl. OECD 2000, S. 10). Der Begriff Literalität wird demnach definiert als “the ability to understand and employ printed information in daily activities, at home, at work and in the community – to achieve one’s goals, and to develop one’s knowledge and potential” (OECD 2000, S. 10).

Literalität ist also die Fähigkeit, verschriftlichte Informationen zu verstehen und im Rahmen von Alltagsaktivitäten – sei es zu Hause, auf der Arbeit und bei gesellschaftlichen Aktivitäten – verwenden zu können, um seine Ziele zu erreichen und um sein Wissen und sein Potential weiterentwickeln zu

können (vgl. OECD 2000, S. 10). Die Literalität gliedert sich in drei Bereiche auf: *Prose literacy* (Umgang mit kontinuierlichen Texten); *Document literacy* (Umgang mit nicht-kontinuierlichen Texten) und *Quantitative literacy* (Umgang mit Zahlen). Somit lässt sich die Literalität Erwachsener durch die Erfassung der Fähigkeiten in diesen drei Bereichen messen (vgl. OECD 2000, S. 10). In Tabelle 9 werden die drei Bereiche im Überblick dargestellt.

Tabelle 9: Drei Bereiche der Literalität (nach OECD 2000, S. 10)

Three Domains of Literacy Skills	
Prose literacy	The knowledge and skills needed to understand and use information from texts including editorials, news stories, brochures and instruction manuals.
Document literacy	The knowledge and skills required to locate and use information contained in various formats, including job applications, payroll forms, transportation schedules, maps, tables and charts.
Quantitative literacy	The knowledge and skills required to apply arithmetic operations, either alone or sequentially, to numbers embedded in printed materials, such as balancing a chequebook, figuring out a tip, completing an order form or determining the amount of interest on a loan from an advertisement.

Messung der Literalität: Im Rahmen der IALS-Studie wurde eine komplexe Methodik angewandt, um die Fähigkeiten einer Person in den jeweiligen Literalitäts-Bereichen erfassen und messen zu können. Aufgrund dieser methodischen Herangehensweise lassen sich die Fähigkeiten Erwachsener in den jeweiligen Bereichen auf einer Skala von 0 bis 500 Punkten abbilden (vgl. OECD 2000, S. 10). Die Fähigkeit einer Person innerhalb eines Literalitäts-Bereiches wird durch einen Wert ausgedrückt. Dieser gibt an, ab wann eine Person mit 80-prozentiger Erfolgsquote die Items des Tests, welche unterschiedliche Schwierigkeitsgrade aufweisen, lösen kann (vgl. OECD 2000, S. 10 f.). Mithilfe der Werte lassen sich unterschiedliche Stufen der Fähigkeit einer Person innerhalb eines Literalitäts-Bereiches definieren (vgl. OECD 2000, S. 11 f.).

Im Folgenden werden die fünf Stufen der Literalität in Bezug auf den Bereich der *Quantitative Literacy* erläutert.

Tabelle 10: Literalitätsstufen des Bereiches Quantitative Literacy (modifiziert nach Jordan 2011a, S. 96)

Literalitätsstufe	Beschreibung der Literalitätsstufe
Stufe 1	Obwohl keine der im Rahmen von IALS verwendeten Aufgaben der Stufe 1 zugeordnet werden kann, wird aus Erfahrung angenommen, dass solche Aufgaben vom Leser eine einzige, relativ einfache Rechenoperation (normalerweise Addition) verlangen, für die entweder die Zahlen auf dem Testbogen schon eingetragen sind und die Rechenoperation vorgeschrieben ist, oder die Zahlen angegeben sind, und die Rechenoperation ohne das Hinzuziehen weiterer Informationen durchzuführen ist.
Stufe 2	Aufgaben in dieser Stufe verlangen im Allgemeinen vom Leser, eine einzige arithmetische Operation (häufig Addition oder Subtraktion) durchzuführen und dabei Zahlen zu benutzen, die im Text oder im Dokument gut zu finden sind. Die erforderliche Rechenoperation kann leicht aus dem Wortlaut der Frage oder dem Aufbau des Materials (z. B. einem Kontoauszug oder einem Bestellformular) abgeleitet werden.
Stufe 3	Die Aufgaben dieser Stufe erwarten vom Leser normalerweise die Durchführung einer einzigen Rechenoperation. Die Operationen variieren jedoch stärker – es finden sich einige Multiplikations- und Divisionsaufgaben in dieser Stufe. Manchmal sind zwei oder mehr Zahlen nötig, um die Aufgabe zu lösen, und die Zahlen sind häufig in komplexe Zusammenhänge eingebettet. Einige Aufgaben verlangen vom Leser schwierige Schlussfolgerungen, um die passende Operation zu bestimmen, obwohl oft ausdrückliche Hinweise im Aufgabentext enthalten sind. Diese Hinweise können eindeutige Fragewörter („Wie viel?“) oder Angaben der erforderlichen Operation („Rechnen Sie die Differenz aus!“) sein, die als semantische Relationsausdrücke bezeichnet werden.
Stufe 4	Mit einer Ausnahme verlangen die Aufgaben dieser Stufe vom Leser die Durchführung einer einzigen Rechenoperation, bei der gewöhnlich entweder die Quantitäten oder die Rechenoperationen nicht leicht zu ermitteln sind. Das heißt, dass bei den meisten Aufgaben dieser Stufe in der Frage oder Anleitung keine semantischen Relationsausdrücke enthalten sind, um den Leser anzuleiten.
Stufe 5	Diese Aufgaben verlangen vom Leser, unterschiedliche Operationen in Folge durchzuführen. Sie müssen die wesentlichen Bestandteile der Rechnung aus dem gelieferten Material oder mithilfe von Hintergrundwissen die benötigten Zahlen und Rechenoperationen finden.

Quantitative Literacy im Kontext schulischer und beruflicher Bildung: Die Gymnasien in Nordrhein-Westfalen haben bereits im Jahr 2007 im Mathematik-Kernlehrplan für das Gymnasium (G8) die Förderung der *Quantitative*

Literacy festgelegt. Hierbei wird hervorgehoben, dass die Schüler am Ende der Sekundarstufe I in der Lage sein sollten, Texten mathematische Informationen entnehmen zu können (vgl. Leuders 2007, S. 13). An dieser Hürde scheitern aber viele der Schüler in der Altenpflegeausbildung. In der beruflichen Bildung haben die Schüler es überwiegend mit Fachtexten zu tun. Diese unterscheiden sich in ihrem inhaltlichen und strukturellen Aufbau wesentlich von den literarischen Texten, welche im Deutschunterricht der allgemeinbildenden Schulen verwendet werden (vgl. Mitterhuber 2008, S. 19 ff.).

4.2 Textaufgaben als Aufgabentyp zur Erfassung von mathematischem Textverständnis

Das mathematische Textverständnis wird in der vorliegenden Arbeit mithilfe von Textaufgaben erfasst. Textaufgaben gelten als einer von mehreren Aufgabentypen des *Sachrechnens* (vgl. Greefrath 2010, S. 84 f.). Auf den genannten Aufgabentyp soll zu einem späteren Zeitpunkt in diesem Kapitel eingegangen werden. Zuvor ist es wichtig, sich semantisch dem Begriff *Sachrechnen* zu nähern. Greefrath (2010, S. 12) bezeichnet Sachrechnen im weitesten Sinne als „die Auseinandersetzung mit der Umwelt sowie die Beschäftigung mit wirklichkeitsbezogenen Aufgaben...“. Um manche Aufgabentypen des Sachrechnens bearbeiten zu können, wird auf das mathematische Modellieren zurückgegriffen. „Für das mathematische Modellieren ist ... immer eine Situation aus der Realität der Ausgangspunkt. Die Situation wird dann mit Hilfe eines mathematischen Modells beschrieben und bearbeitet“ (Greefrath 2010, S. 44). Dieses Vorgehen wird als Modellieren bezeichnet (vgl. Greefrath 2010, S. 44). Dies gilt jedoch nicht für jeden Aufgabentyp des Sachrechnens wie etwa den der Textaufgaben. „Textaufgaben sind typisch für das klassische Sachrechnen. Sie bestehen aus Aufgaben in Textform – teilweise auch ergänzt durch ein Bild. Die Sache ist ... austauschbar, und die Realität ist häufig sehr vereinfacht dargestellt. Das Ziel ist die Förderung mathematischer Fähigkeiten“ (Greefrath 2010, S. 84). Bei *Textaufgaben* handelt es sich zudem „um Rechenaufgaben oder mathema-

tische Aufgaben ..., die sich auf Sachen, also Sachverhalte, beziehen“ (Fricke & Besuden 1987, S. 5). Im Rahmen von Textaufgaben werden Themengebiete wie Stückzahlen, Geldbeträge, Längen, Zeitspannen, Gewichte, Temperaturen sowie Flächen- und Rauminhalte bearbeitet. Zudem können Sachrechenaufgaben auch Elemente aus der Statistik und Größenlehre beinhalten (vgl. Winter 2003, S. 15).

Textaufgaben eignen sich nicht primär, um ein mathematisches Modell zu erstellen. Durch den fehlenden Realitätsbezug und die vorgegebene Vereinfachung wird bei der Bearbeitung von Textaufgaben eher von einer Mathematisierung gesprochen (vgl. Greefrath 2010, S. 84 f.). Mathematisieren ist eine Teilkompetenz des Modellierens und umschreibt die Fähigkeit, Situationen in mathematische Modelle übersetzen zu können (vgl. Greefrath 2010, S. 52).

Neben den Textaufgaben gibt es noch weitere Aufgabentypen, wie beispielsweise eingekleidete Aufgaben und Sachaufgaben respektive Sachprobleme (vgl. Maier 1976, S. 161; Fricke & Besuden 1987, S. 7; Krauthausen & Scherer 2010, S. 83 ff.; Greefrath 2010, S. 83 ff.), welche hier kurz erläutert werden.

Eingekleidete Aufgaben sind dadurch charakterisiert, dass sie sich unabhängig vom Textinhalt der Aufgabe ausschließlich auf die Verbindung von Zahlen und Rechenoperationen konzentrieren (vgl. Fricke & Besuden 1987, S. 7; Maier 1976, S. 161). Sie haben also einen lebensnahen beziehungsweise modellhaften Charakter. Sie sollen den Schülern alltägliche mathematische Problemlagen aufzeigen, die mittels mathematischer Rechenoperationen zu lösen sind (vgl. Maier 1976, S. 161 ff.). Beispiel: „Hans hat sieben Kastanien. Er findet noch fünf dazu.“ (Maier 1976, S. 161)

Neben dem eingekleideten Aufgabentyp existiert ein weiterer Aufgabentyp: die *Sachprobleme*. Im Fokus der Sachprobleme stehen keine Rechenoperationen, sondern das Lösen von lebensnahen Sachproblemen. Maier konstatiert, dass für das Lösen dieser Sachprobleme rechnerische Fähigkeiten

notwendig sind (vgl. Maier 1976, S. 164). Beispiel: „Was leistet das menschliche Herz?“ (Maier 1976, S. 164). Tabelle 11 zeigt zusammenfassend die unterschiedlichen Aufgabentypen mit deren Zielsetzungen, Sinn und Kontexten im Vergleich auf.

Tabelle 11: Charakteristika der jeweiligen Aufgabentypen (nach Maier 1976, S. 166)

Aufgabentyp:	Eingekleidete Aufgabe	Textaufgabe	Sachprobleme
Ziel:	Lösen von Rechenoperationen und Anwenden von Rechenverfahren	Quantifizierung von Sachverhalten	Lösen von Sachproblemen und mittelbare Informationsgewinnung
Sinn:	Erwerben rechnerischer und mathematischer Begriffe und Einsichten	Verstehen des Zahl-Sache-Zusammenhangs	Erwerben der Fähigkeit und dauernden Bereitschaft zur rechnerischen Daseinserhellung und Lebensbewältigung
Zusammenhang:	Zahlen absolut wichtig: Sachzusammenhang belanglos und auswechselbar	Ausgewogenes Verhältnis von Zahl und Sache	Sachzusammenhang wichtig: Zahlen veränderlich und relativ bedeutsam

4.3 Strategien zum Lösen und Verstehen von Textaufgaben

Beim Lösen einer Textaufgabe muss zuerst der Textinhalt der Aufgabe erschlossen und danach die in der Textaufgabe befindlichen Zusammenhänge verknüpft werden. Im Verlauf dieses Prozesses oder gar nach dessen Beendigung gelingt es dem Schüler, die in der Textaufgabe befindliche mathematische Aufgabe zu lösen (vgl. Reusser 1997, S. 149). Reusser versteht den Prozess zur Lösung einer Textaufgabe als „den von einer (expliziten oder impliziten) Problemfrage geleiteten planvollen Vorgang, bei welchem *ausgehend* von einem Problemtext eine sachhaltige Situationsvorstellung – also ein auf die Handlungs- und Prozeßstruktur der Aufgabe bezogenes mentales Situationsmodell – gebildet und schrittweise auf ein quantitatives Gerüst – das mathematische Problemmodell – reduziert wird“ (Reusser 1997, S. 153). Die folgende Abbildung 3 zeigt den Prozess, welcher bei der Bearbeitung einer Textaufgabe durchlaufen werden kann.

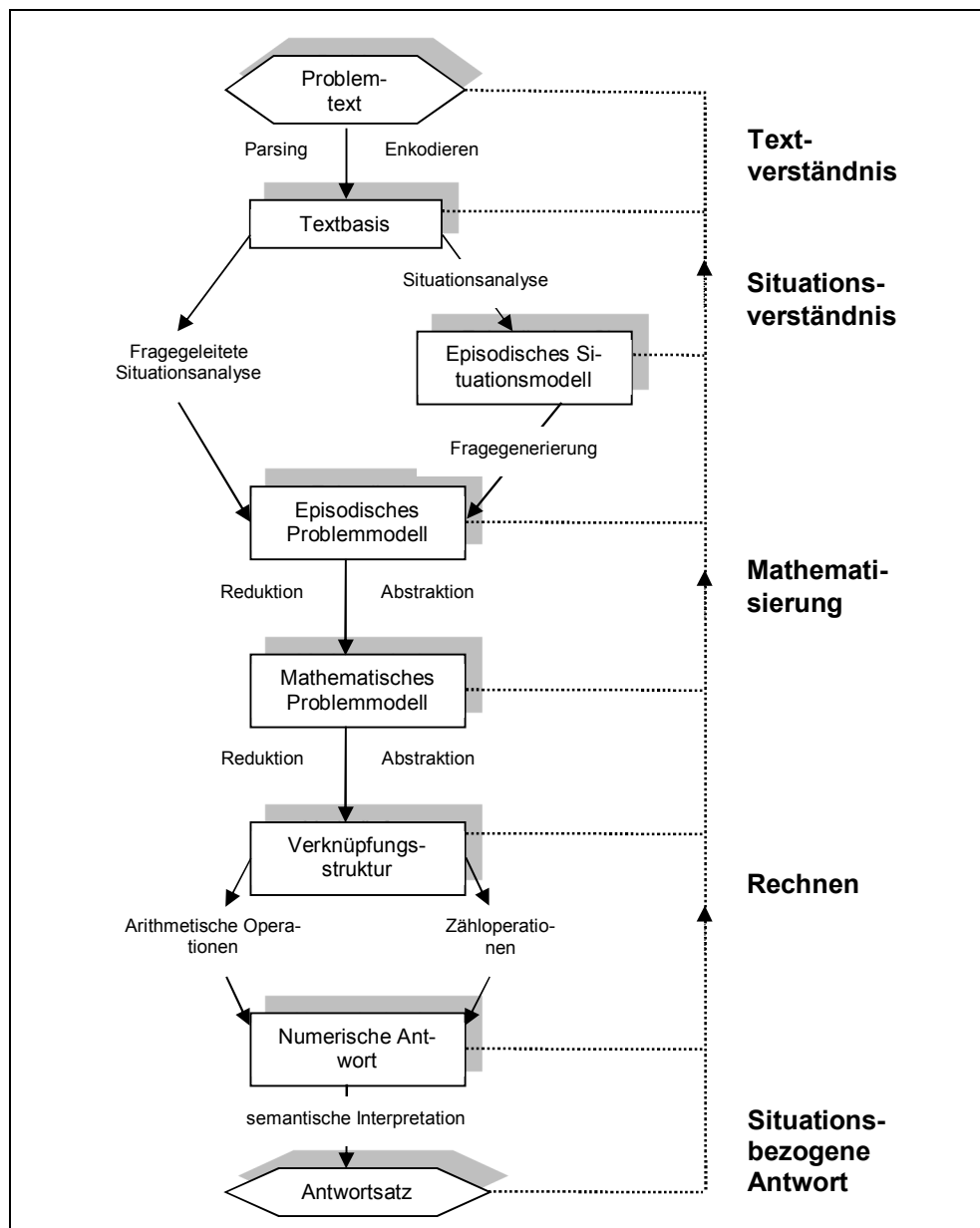


Abbildung 3: Situation-Problem-Solver-Modell (nach Reusser 1992, S. 231)

Das Situation-Problem-Solver-Modell (SPS-Modell) umfasst fünf Phasen: die Phase des Textverständnisses, die Phase des Situationsverständnisses, die Phase der Mathematisierung, die Phase des Rechnens und die Phase der Erstellung einer situationsbezogenen Antwort. In der Phase des Textverständnisses versucht der Schüler, den Aufgabentext zu verstehen und eine Textbasis zu erstellen. Die Textbasis dient dazu, ein episodisches Situations- oder Problemmodell zu entwickeln. Hierbei nutzt der Schüler sein allgemeines und aufgabenspezifisches Wissen. In diesem Schritt ar-

beitet er die in dem Aufgabentext sich befindende Handlungs- oder Problemsituation heraus. Anschließend versucht der Schüler, die Handlungs- oder Problemsituation noch weiter zu konkretisieren, und bildet das für die Situation passende mathematische Problemmodell, bevor er daraus eine mathematische Gleichung oder Rechnung ableiten kann. Als Ergebnis erhält der Schüler ein numerisches Ergebnis. Im letzten Schritt hat er die Aufgabe, das numerische Ergebnis in einen zur Handlungs- oder Problemsituation passenden Antwortsatz zu übertragen (vgl. Reusser 1992, S. 230 f.).

Das Prozessmodell nach Reusser basiert in Bezug auf die Erstellung einer Textbasis aus einem Problemtext bis zu dem entsprechenden Situations- oder Problemmodell auf den Ausführungen von Dijk und Kintsch (1983) zur Theorie einer strategischen Textverarbeitung (vgl. Reusser 1992, S. 229; Reusser 1997, S. 149). Im Folgenden soll näher auf die Theorie von Dijk und Kintsch in Verbindung mit dem Prozessmodell nach Reusser eingegangen werden.

Damit ein Schüler von einem Problemtext zu einem Situations- oder Problemmodell gelangt, bedarf es einiger Zwischenschritte oder auch „Repräsentationsebenen“ (Kintsch 1994, S. 41) genannt. Eine *Repräsentation* ist eine Bedeutungsvorstellung. Diese wird benötigt, um die Aussage eines Problemtextes darstellen zu können (vgl. Kintsch 1994, S. 41). Es gibt verschiedene Repräsentationsebenen, wie beispielsweise die Oberflächenstruktur und die Textbasis. Die Oberflächenstruktur dient der Darstellung der Wörter und Sätze eines Textes und ihrer linguistischen Struktur. Die Textbasis stellt den semantischen Gehalt und die Organisation eines Textes dar. Aus der Oberflächenstruktur eines Textes – also aus den einzelnen Wörtern und Sätzen – werden *Bedeutungseinheiten* respektive *Propositionen* (vgl. Kintsch 1994, S. 41). Eine Proposition setzt sich aus einem Prädikat und aus einem oder mehreren Argumenten zusammen. Das Prädikat dient dazu, die Argumente miteinander in Beziehung setzen zu können (vgl. van Dijk & Kintsch 1983, S. 113). Folgendes Beispiel soll dies verdeutlichen:

Proposition Nr.1 (P1): PFLEGEN(PFLEGEFACHPERSON,BEWOHNER)

In dem genannten Beispiel stellt PFLEGEN das Prädikat dar und bringt die beiden Argumente PFLEGEFACHPERSON und BEWOHNER miteinander in Verbindung respektive in Beziehung. Wenn P1 als Satz ausformuliert wird, dann lautet dieser Satz: *Pflegefachperson pflegt Bewohner*. Prädikate können oder müssen nicht immer über Verben definiert werden. Auch Argumente können nicht immer durch Substantive oder Nomen gebildet werden (vgl. van Dijk & Kintsch 1983, S. 113), wie das folgende Beispiel zeigt: Aus dem Satz: *Wenn ein Bewohner Durst hat, muss er trinken*. Aus diesem Satz lassen sich folgende Propositionen ableiten:

Proposition Nr. 2 (P2): KAUSALITÄT (WENN(P3,P4))

Proposition Nr. 3 (P3): HABEN(BEWOHNER,DURST)

Proposition Nr. 4 (P4): MÜSSEN(BEWOHNER, P5)

Proposition Nr. 5 (P5): TRINKEN(BEWOHNER)

P2 verbindet die beiden Propositionen P3 und P4 in einer Art kausalen Beziehung. Um eine Kausalität zwischen einer und mehreren Propositionen beschreiben zu können, ist es wichtig zu konstatieren, dass Verstehen „als Interaktion und Verschmelzung des Textes mit dem aktualisierten Wissen bzw. dem persönlichen Erfahrungsschatz des Verstehenden gesehen“ (Kintsch 1994, S. 43) wird. Im aufgeführten Beispiel entstehen mehrere kleine Vernetzungen zwischen den Propositionen – auch *Mikrostruktur* genannt. Die Mikrostruktur stellt die Textbasis eines Textes dar. Die Propositionen innerhalb der Mikrostruktur sind nach ihrer Bedeutsamkeit geordnet. Dies ist wichtig, da sich Schüler bedeutsame Propositionen besser merken und dadurch die dazugehörige Mikrostruktur im Langzeitgedächtnis eher abspeichern können (vgl. Kintsch 1994, S. 42). Bei dem Beispiel handelt es sich um einen kleinen Satz beziehungsweise kleinen Textumfang. Aufgrund der Tatsache, dass die Kapazität des menschlichen Arbeitsgedächtnisses begrenzt ist (vgl. Kintsch 1994, S. 42), können Schüler nur begrenzte Mengen an Textumfang auf einmal verarbeiten beziehungsweise einen kleinen Text in eine propositionale Bedeutungsrepräsentation umwandeln. Ein umfangreicherer Text wird daher zuerst fragmentiert bearbeitet. Es werden zunächst mehrere Mikrostrukturen nacheinander gebildet. Um die einzelnen

Mikrostrukturen kohärent miteinander zu verbinden, werden die wichtigsten übergeordneten Propositionen der einzelnen Mikrostrukturen im Kurzzeitgedächtnis gespeichert. Somit besteht die Möglichkeit, eine propositionale Bedeutungsrepräsentation für einen umfangreicheren Text zu generieren. Es entsteht eine *Makrostruktur* (vgl. Kintsch 1994, S. 42). Kintsch beschreibt diesen Prozess im Konstruktions-Integrations-Modell. Innerhalb der Konstruktionsphase wird, wie bereits beschrieben, ein Netzwerk von propositionalen Bedeutungsrepräsentationen erstellt, die durch übergeordnete Propositionen mehr oder weniger kohärent miteinander verbunden sind. Diese Verbindungen können aber noch sehr chaotisch und nur für den einzelnen Schüler verständlich konstruiert worden sein (vgl. Kintsch 1994, S. 43; Kintsch 1988). Erst in der Integrationsphase versucht der Schüler mithilfe seines vorhandenen (Welt-)Wissens und seines Erfahrungsschatzes, wichtige von unwichtigen Informationen auszusortieren und übergeordnete Propositionen so zu verändern, dass die einzelnen Mikrostrukturen in eine sinnvolle und plausible Makrostruktur überführt werden können. Es entsteht ein *episodisches Situationsmodell* (vgl. Reusser 1992). Zur Lösung einer Textaufgabe entwickelt der Schüler zusätzlich ein *episodisches Problemmodell* (vgl. Reusser 1992), welches die mathematischen Operationen repräsentiert, die zur Lösung der Aufgabe benötigt werden können. Die einzelnen Prozesse laufen nicht nacheinander ab, sondern können parallel zueinander verlaufen (vgl. Kintsch 1994, S. 43; Kintsch 1988).

4.4 Zusammenfassende Betrachtung des vierten Kapitels

Kapitel 4 befasste sich unter anderem mit der definitorischen Bestimmung des zu untersuchenden theoretischen Konstruktes. Unter mathematischem Textverständnis wird, kurz gesagt, die Fähigkeit verstanden, aus einer schriftlichen Information die relevanten Informationen von den irrelevanten zu trennen, um mit den relevanten weiterarbeiten beziehungsweise mit den Daten und Zahlen eine oder mehrere Rechenoperationen durchführen zu können. Zur Erfassung des mathematischen Textverständnisses wurde als Aufgabentyp die Textaufgabe gewählt. Das Lösen solch mathematischer

Textaufgaben stellt einen Prozess dar, welcher von Reusser (1992) detailliert erläutert wird. Das Situation-Problem-Solver-Modell (SPS-Modell) dient zur differenzierten Beschreibung des Lösungsweges in Bezug auf eine Textaufgabe.

II. Methodischer Teil

Im methodischen Teil der Arbeit wird in Kapitel 5 die methodische Herangehensweise der Untersuchung beschrieben, bevor in Kapitel 6 für die Auswertung der Daten wichtige Testtheorie und deren Testmodelle erläutert werden. Abschließend findet in Kapitel 7 eine ausführliche Darstellung und Beschreibung des Tests zur Erfassung des mathematischen Textverständnisses (TeMaTex) statt.

5 Methoden und Vorgehensweisen der Studie

In Kapitel 5.1 sollen die ethischen Überlegungen vorgestellt werden, bevor in Kapitel 5.2 auf die Berechnung der Stichprobengrößen eingegangen wird. Die Kapitel 5.3 und 5.4 dienen der Beschreibung in Bezug auf die Durchführung der Datenerhebung und auf die Datenauswertung. In Kapitel 5.5 wird die methodische Herangehensweise einer Expertenbefragung erläutert. Kapitel 5.6 stellt die methodischen Grundlagen zur Bestimmung der Gütekriterien von Testverfahren dar. Abschließend findet in Kapitel 5.7 eine Methodendiskussion und in Kapitel 5.8 eine Zusammenfassung des gesamten Kapitels statt.

5.1 Ethische Überlegungen und Datenschutz

Im Rahmen der Studie soll das mathematische Textverständnis von Altenpflegeschülern zu Beginn der dreijährigen Altenpflegeausbildung erfasst werden. Das theoretische Konstrukt des mathematischen Textverständnisses ist ein personenorientiertes Merkmal. Um dieses Merkmal erfassen zu können, werden die Probanden schriftlich befragt. Der Test zur Erfassung des mathematischen Textverständnisses umfasst ausschließlich mathematische Aufgabenstellungen. Es gibt keinen Hinweis darauf, dass die Aufgaben die körperliche und seelische Integrität der Befragten beeinflussen könnten. Daher werden keine vorbeugenden oder nachsorgenden Maßnahmen zum Schutz der Probanden geplant. Auf eine schriftliche Einverständniserklärung wird auch verzichtet. Stattdessen wird eine mündliche Einver-

ständniserklärung abgegeben. Die Voraussetzung für eine informierte Zustimmung ist die Information der Probanden. Dies erfolgt über den Lehrervortrag unmittelbar vor der Befragung. Nach der Information über die Studie werden die Probanden mündlich nach deren Einwilligung zur Teilnahme an der Studie befragt. Die erhobenen Daten der Probanden werden ausschließlich auf einem Datenträger gespeichert, auf den nur der Forscher einen Zugriff hat. Die Daten unterliegen den Vorgaben des Datenschutzgesetzes. Zur statistischen Auswertung werden die Daten codiert. Nach der Codierung enthalten diese keinen Hinweis mehr auf die persönliche Identität der Probanden und sind somit anonymisiert. Es wird hiermit versichert, dass die Daten ausschließlich für die vorliegende wissenschaftliche Untersuchung verwendet werden.

5.2 Stichprobengrößen

Die Berechnung der erwünschten Stichprobengrößen für die Studie basiert im Grundsatz auf der möglichen maximalen Anzahl von Antwortmustern (patterns) für die einzelnen Itemkonstellationen. Dies bedeutet, dass jedes der möglichen Antwortmuster zumindest theoretisch die Möglichkeit haben sollte, in den Stichproben enthalten zu sein. In der vorliegenden Studie gibt es – aus dem Test zur Erfassung des mathematischen Textverständnisses (TeMaTex) – insgesamt 18 Items. Für die Version in Bezug auf das Expertenrating sind es insgesamt 17 Items. Da der TeMaTex und die überarbeitete Version für das Expertenrating ausschließlich Aufgaben enthalten, die dichotomisiert ausgewertet werden können, was heißt, dass für jede richtige Antwort ein Punkt und für jede falsche Antwort kein Punkt erteilt wird, gibt es $2^{18} = 262.144$ respektive $2^{17} = 131.072$ mögliche Antwortmuster. Daraus folgt, dass die Stichproben unrealistisch groß sein müssten, damit alle diese Ergebnisse realisiert werden können. Aus diesen Gründen wird im Rahmen von χ^2 -basierten Modellgeltungstests auf das Bootstrap-Verfahren verwiesen, das in Kapitel 6.1.5.3.3 vorgestellt wird (vgl. Rost 2004). Eid und Schmidt weisen darauf hin, dass unter Verwendung des einparametrischen

Rasch-Modells eine Stichprobengröße von 100 bis 200 Personen ausreichend wäre, um stabile Schätzungen durchführen zu können (vgl. Eid & Schmidt 2014, S. 213 f.). Bühner empfiehlt eine Stichprobengröße zwischen 250 und 300 Personen (vgl. Bühner 2011, S. 583). Insgesamt haben sich 411 Probanden in der ersten Erhebungsphase und 253 in der zweiten Erhebungsphase zur Teilnahme an der Studie bereit erklärt.

5.3 Beschreibung der Datenerhebungen

Die erste Datenerhebung zur Überprüfung der Konstruktvalidität des TeMaTex und zur Erfassung des mathematischen Textverständnisses fand in dem Zeitraum vom 01.09.2013 bis 01.04.2014 statt, an der 12 Altenpflegeschulen aus verschiedenen Bundesländern der Bundesrepublik Deutschland beteiligt waren. Die Auszubildenden befanden sich zum Zeitpunkt der Datenerhebung im ersten Theorieblock zu Beginn des ersten Lehrjahres in der Altenpflegeschule. Die eigentliche Durchführung des Tests fand in einer Unterrichtsstunde von 45 Minuten eines vorher vereinbarten Schultages statt. Der Testleiter war der Lehrgangsverantwortliche (Klassenlehrer) des Altenpflegekurses. Der Testleiter informierte die Probanden vorab über das Ziel, den Zweck und den Ablauf der Untersuchung im Rahmen eines kurzen Lehrervortrags. Zusätzlich holte er die mündliche Einverständniserklärung der Probanden ein. Probanden, die ihre Einwilligung verweigerten, wurden bei der Datenerhebung nicht berücksichtigt. Nach dem Klären von Rückfragen teilte der Testleiter die Testhefte mit den Bearbeitungsbögen aus. Das Testheft umfasst zusätzlich Informationen, die das Prozedere hinsichtlich der Bearbeitung des Testheftes genau erläutern. So konnte jeder Proband den Ablauf der Untersuchung für sich erneut nachvollziehen. Es durften zur Bearbeitung der Aufgabenstellungen keine Hilfsmittel, wie beispielsweise ein Taschenrechner, verwendet werden. Für den Test wurden 45 Minuten Bearbeitungszeit inklusive Testinstruktion vorgegeben. Abschließend wurden die Testhefte und Bearbeitungsbögen wieder vom Testleiter eingesammelt. Ergänzend wurden die Probanden darüber informiert, dass sie die Ergebnisse der Befragung gerne zur Verfügung gestellt bekämen.

Die zweite Datenerhebung zur Validierung des CareMaTex fand in dem Zeitraum vom 01.03.2016 bis 01.07.2016 statt. An der zweiten Erhebungsphase beteiligten sich vier Altenpflegeschulen aus verschiedenen Bundesländern der Bundesrepublik Deutschland. Die Auszubildenden befanden sich zum Zeitpunkt der Datenerhebung im ersten bis dritten Lehrjahr der Altenpflegeausbildung. Das weitere Prozedere der zweiten Datenerhebung gleicht der ersten Erhebungsphase und wird daher nicht weiter dargestellt.

5.4 Durchführung der Datenauswertungen

Im ersten Schritt wurde mithilfe des package Rcmdr des Statistikprogramms R eine deskriptive Analyse der erhobenen Daten durchgeführt, in deren Verlauf die Häufigkeit in Bezug auf eine Variable ermittelt wurde. Zunächst wurden fehlende Angaben der Probanden bereinigt. Danach konnte mit der Datenauswertung begonnen werden. Die Auswertung der Daten erfolgte auf zwei unterschiedlichen Analyseebenen, einerseits auf der Ebene der beschreibenden Datenanalyse – dazu gehörte zum Beispiel die Darstellung der zentralen Tendenz zur Beschreibung der Stichprobe –, andererseits auf der Ebene der schließenden Datenanalyse zur Überprüfung der in Kapitel 1.2 formulierten Hypothesen durch geeignete Testverfahren, wie beispielsweise die Rasch-Analyse (vgl. LoBiondo-Wood & Haber 1996).

Die deskriptive Ebene der Datenanalyse dient, wie bereits erwähnt, zur Beschreibung der Stichproben und darüber hinaus zur Beschreibung der Häufigkeitsverteilung in Bezug auf die Antworten der Probanden. Von besonderem Interesse bei der Beschreibung der Stichproben sind die Variablen Alter und Schulabschluss. Die Variable Alter hat einen Intervall-Charakter, die Variable Schulabschluss ist ordinaler Natur. Aufgrund des Intervall-Charakters der Variable Alter wird zur Beschreibung der Stichprobe im Rahmen der ersten Erhebungsphase der Median zur Bestimmung der zentralen Tendenz gewählt. Der Median wird verwendet, um eine Verteilung in zwei Hälften einzuteilen. In jeder Hälfte befinden sich dieselbe Anzahl an Messwerten. Zur Berechnung des Medians werden zunächst alle Messwerte der Größe nach geordnet. Der Messwert, der in der Mitte steht, wird als Median

bezeichnet. Bei einer ungeraden Anzahl von Messwerten wie 2, 4, 7, 8, 9, 10, 12 ist der Median $Md = 8$. Bei einer geraden Anzahl von Messwerten wie 2, 4, 7, 8 ist der Median die numerische Mitte der beiden mittleren Zahlen 4 und 7, also $Md = 5,5$. Zur Berechnung des Median wird mindestens Ordinalskalenniveau vorausgesetzt (vgl. Rasch et al. 2006, S. 16). Für die ordinale Variable Schulabschluss wird die Maßzahl der prozentualen Verteilung zur Bestimmung der Werte verwendet. Im Rahmen der Berechnungen zur Häufigkeitsverteilung in Bezug auf die Antworten der Probanden dienen Maßzahlen wie die Prozentangabe und die Häufigkeit (vgl. LoBiondo-Wood & Haber 1996).

Im zweiten Schritt wurden die Testergebnisse aus den Erhebungen in ein eigens für die Auswertung der TeMaTex-Erhebungsbögen entwickeltes Excel-Programm eingegeben. Nach Eingabe werden die Ergebnisse der Testaufgaben in eine dichotome Datenmatrix umcodiert, das heißt, dass alle Testaufgaben, die korrekt gelöst wurden, mit 1 kodiert werden. Alle nicht gelösten und nicht bearbeiteten Testaufgaben werden mit 0 kodiert. Durch dieses Verfahren entsteht eine Rohdatenmatrix, mit der eine Überprüfung in Bezug auf eine Rasch-Konformität durchgeführt und die individuellen Leistungsscores bestimmt werden können. Zur Durchführung der Rasch-Analyse wurden die Programme eRm und WINMIRA sowie das Programm RUMM 2030 eingesetzt. Im Folgenden werden die statistischen Verfahrensweisen derjenigen Programme erläutert, welche in dieser Arbeit Anwendung fanden.

WINMIRA: Die Statistiksoftware WINMIRA in der Version 2001.0012 ist ein Windowsbasiertes Programm. Die Software bietet sich zur Berechnung von Rasch-Modellen an (vgl. von Davier 2001, S. 8). Bei WINMIRA werden die Itemparameter mit der *conditional maximum likelihood* (cML) geschätzt (vgl. Rost 2004, S. 311). Die Personenparameterschätzungen werden mit der *unconditional maximum likelihood* (uML) und mit der *weighted likelihood estimation* (WLE) durchgeführt (vgl. von Davier 2001, S. 70 f.). Für die Angabe der Item-Fit-Werte werden der ZQ-Wert und die Q-Indizes genutzt

(vgl. von Davier 2001, S. 77). Im Rahmen der Überprüfung der Modellgeltung werden unter anderen die Likelihoods sowie verschiedene χ^2 -Statistiken wie Pearsons χ^2 und Cressie Read χ^2 angeboten (vgl. von Davier 2001, S. 22 f.). Mit WINMIRA kann auch ein parametrischer Bootstrap durchgeführt werden (vgl. von Davier 2001, S. 33 ff.).

RUMM 2030: Die Statistiksoftware RUMM 2030 (Rasch Unidimensional Measurement Models) in der Version 2030 ist wiederum ein windowsbasiertes Programm. Die Software bietet sich ausschließlich zur Berechnung von latent-trait-bezogenen Rasch-Modellen an. Bei RUMM 2030 werden die Itemparameter mit der Pairwise-Methode geschätzt (vgl. Rost 2004, S. 311). Für die Angabe der Item-Fit-Werte werden die Fit-Residuals und die entsprechenden χ^2 -Statistiken der Items verwendet. Im Rahmen der Überprüfung der Modellgeltung wird eine χ^2 -Statistik angeboten (vgl. Andrich et al. 2009, S. 19 ff.). Zudem bietet RUMM 2030 die Berechnung des Differential Item Functioning und die Berechnung des Andrich-Reliabilitätskoeffizienten zur Angabe der internen Konsistenz an (vgl. Andrich et. al. 2009).

eRm: Die Anwendung *Extended Rasch modeling* (eRm) in der Version 0.10-2 ist ein Ergänzungsmodul (package) der Statistiksoftware R in der Version 2.9.1. R und eRm sind kostenfrei im Internet verfügbar. Bei eRm werden die Itemparameter über die CML-Methode berechnet, die Personenparameter über die UML und mit der WLE-Methode (vgl. Mair & Hatzinger 2009a, S. 6 f.). Für die Angabe der Item-Fit-Werte werden die Fit-Residuals und die Ergebnisse aus den Wald-Statistiken angeboten (vgl. Mair & Hatzinger 2009b, Mair & Hatzinger 2009a, S. 11). Im Rahmen der Überprüfung der Modellgeltung werden der Andersen-LR-Test und ein grafischer Modelltest angeboten (vgl. Mair & Hatzinger 2009b, Mair & Hatzinger 2009a).

5.5 Expertenbefragung

Von Mitte Januar 2016 bis Mitte Februar 2016 wurde ein Expertenrating durchgeführt, da die Items des TeMaTex an den Kontext der Altenpflege angepasst werden mussten. Hierbei wurde die Struktur der Aufgaben aus

dem TeMaTex übernommen. Lediglich der Inhalt der Texte und der Aufgaben wurde verändert. Um die inhaltliche Validität der Items des CareMaTex einschätzen zu können, wurden die verbleibenden 17 Items bundesweit verschiedenen Experten aus dem pädagogischen Bereich der Altenpflege vorgelegt. Sie verfügen über Erfahrungen in dem entsprechenden Bereich der Altenpflegeausbildung. Insgesamt wurden 12 Altenpflegeschulen aus dem privaten und öffentlichen Schulsektor angeschrieben. Zu den Experten zählten ausschließlich Schulleiter der entsprechenden Einrichtungen. Die Rücklaufquote betrug 58 % – also nahmen 7 von 12 Experten der angeschriebenen Altenpflegeschulen an dem Expertenrating teil.

In einem Beurteilungsbogen sollten die Experten einschätzen, inwieweit die Fähigkeiten zur Lösung der jeweiligen Items für den Altenpflegeberuf relevant sind. Im Folgenden soll beispielhaft eine Aufgabe aus dem Beurteilungsbogen dargestellt werden.

Tabelle 12: Aufgabe 1 aus dem Beurteilungsbogen

Evaluationsfrage:	sehr relevant	relevant	wenig relevant	nicht relevant	Kommentar:
Ich halte die zur Lösung von Aufgabe 1 benötigten Fähigkeiten für den Altenpflegeberuf für...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

Die Experten hatten die Möglichkeit, mittels einer Likert-Skalierung die Relevanz aller 17 Items anzugeben. Die Skalierung reichte von *sehr relevant* bis *nicht relevant*. Zu jeder Evaluationsfrage konnten die Experten einen zusätzlichen Kommentar hinzufügen. In einer abschließenden Evaluationsfrage konnte der Test in seiner Gesamtheit bewertet werden – was im Folgenden dargestellt wird.

Tabelle 13: Evaluationsfrage aus dem Beurteilungsbogen

Evaluationsfrage:	trifft zu	trifft eher zu	trifft eher nicht zu	trifft nicht zu	Kommentar:
Der Test in seiner Gesamtheit gibt das Anforderungsprofil im Bereich der Fähigkeit „Verständnis mathemathaltiger Texte“ für den Altenpflegeberuf angemessen wieder:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

Die Skalierung reichte bei dieser Frage von *trifft zu* bis *trifft nicht zu*. Gegen Ende des Beurteilungsbogens konnten die Experten noch Anregungen und Wünsche bezüglich der Items und deren Überarbeitung notieren.

Für die Güte der Übereinstimmung der Rangreihen m , welche die Experten durch die Einschätzung der Items nach ihrer Relevanz gebildet haben, wurde der Konkordanzkoeffizient W nach Kendall und Babington-Smith gewählt. Der Konkordanzkoeffizient W ist ein Maß der Zuverlässigkeit für mindestens ordinalskalierte Ratings (vgl. Wirtz & Caspar 2002) und kann Werte zwischen 0 und 1 annehmen, wobei 0 für eine perfekte Diskordanz (sehr schwache Übereinstimmung) und 1 für eine perfekte Konkordanz (sehr starke Übereinstimmung) steht (vgl. Bortz & Lienert 2003, S. 300). W gibt den Anteil der Varianz zwischen den Ratingrängen der verschiedenen Items an der Gesamtvarianz der Ränge an. Wenn der Koeffizient einen Wert von 1 erreicht, dann liegen dieselben Items bei allen Experten auf den gleichen Ratingrängen (vgl. Wirtz & Caspar 2002, S. 135). Aus folgender Gleichung lässt sich der Konkordanzkoeffizient berechnen (vgl. Bortz & Lienert 2003; Wirtz & Caspar 2002):

$$(Gl. 1) \quad W = \frac{12 * \sum_{i=1}^N (T_i - \bar{T})^2}{m^2 * (N^3 - N)}$$

T_i steht für die Rangsumme des jeweiligen Items. \bar{T} gibt den Durchschnitt aller Rangsummen an. Unter m wird die Anzahl der Experten verstanden. Die Anzahl der Items wird als N gekennzeichnet.

5.6 Bestimmung der Gütekriterien eines Testverfahrens

Die Qualitätsanforderungen an einen Test, welcher unter anderem theoretische Konstrukte erfassen soll, umfassen mehrere Gütekriterien (vgl. Moosbrugger & Kelava 2012a, S. 2 f.). Ein Test muss diesen Gütekriterien Rechnung tragen, „um wissenschaftlichen Ansprüchen zu genügen“ (Moosbrugger & Kelava 2012a, S. 2 f.). Nach Moosbrugger und Kelava (2012a, S. 3) gibt es die folgenden Kriterien: Objektivität, Reliabilität, Validität, Skalierung, Normierung (Eichung), Testökonomie, Nützlichkeit, Zumutbarkeit, Unverfälschbarkeit und Fairness. Von besonderer Bedeutung sind (vgl. Jordan 2011a) die folgenden drei Testgütekriterien: Objektivität, Reliabilität und Validität.

Objektivität: „Ein Test ist dann objektiv, wenn er dasjenige Merkmal, das er misst, unabhängig von Testleiter und Testauswerter misst. Außerdem müssen klare und anwenderunabhängige Regeln für die Ergebnisinterpretation vorliegen“ (Moosbrugger & Kelava 2012b, S. 8). Völlige Objektivität eines Tests läge dann vor, wenn der Test unabhängig vom Testleiter immer gleich durchgeführt und unabhängig vom Testauswerter immer gleich ausgewertet würde (vgl. Moosbrugger & Kelava 2012b, S. 8). Der etwas allgemein gehaltene Begriff Objektivität wird in drei Unterkategorien eingeteilt. Dazu gehören die Durchführungs-, Auswertungs- und Interpretationsobjektivität (vgl. Moosbrugger & Kelava 2012b, S. 8), auf die im folgenden Abschnitt näher eingegangen wird.

Durchführungsobjektivität: Eine hohe Durchführungsobjektivität wird durch „kontrollierte Durchführungsbedingungen“ (Moosbrugger & Kelava 2012b, S. 9) erreicht. Darunter ist zu verstehen, dass die Bedingungen bei der Durchführung des Testverfahrens immer gleich sein müssen. Dies bezieht sich auf das Testmaterial, die Testzeit und auf die Anleitungen, wie die Probanden den Test auszufüllen haben. Es muss auch geregelt sein, wie mit Fragen zum Testverfahren seitens der Probanden umzugehen ist (vgl. Moosbrugger & Kelava 2012b, S. 9). Hierdurch wird eine hohe „Standardisierung“ (Moosbrugger & Kelava 2012b, S.9) der Testdurchführung erreicht.

Vor diesem Hintergrund sei noch erwähnt, dass eine Verschriftlichung der Testinstruktionen einer mündlichen Testinstruktion vorzuziehen ist (vgl. Moosbrugger & Kelava 2012b, S. 9).

Auswertungsobjektivität: Eine hohe Auswertungsobjektivität hingegen hängt von einer einheitlichen Auswertungspraxis der *Testauswerter*, also derjenigen ab, welche die Testergebnisse auswerten (vgl. Moosbrugger & Kelava 2012b, S. 10). Es muss im Testmanual zum Beispiel klar beschrieben sein, welche Antwort als richtig zu bewerten ist (vgl. Moosbrugger & Kelava 2012b, S. 10). Moosbrugger und Kelava bezeichnen dies als „Übereinstimmung verschiedener Testauswerter“ (Moosbrugger & Kelava 2012b, S. 10).

Interpretationsobjektivität: Eine hohe Interpretationsobjektivität liegt vor, wenn unterschiedliche Testauswerter bei denselben vorliegenden Testergebnissen zu gleichen Interpretationen kommen (vgl. Moosbrugger & Kelava 2012b, S. 10). Dies wird durch Angaben von Normtabellen im Testmanual unterstützt. Hierdurch können Testauswerter die individuellen Testergebnisse der Probanden mit der Eichstichprobe vergleichen und daraus Schlussfolgerungen ableiten (vgl. Moosbrugger & Kelava 2012b, S. 10.).

Reliabilität: Moosbrugger und Kelava sprechen bei dem Begriff Reliabilität auch von der „Messgenauigkeit des Tests“ (Moosbrugger & Kelava 2012b, S. 11). Ein Test ist reliabel – also zuverlässig –, „wenn er das Merkmal, das er misst, exakt, d. h. ohne Messfehler, misst.“ (Moosbrugger & Kelava 2012b, S. 11) Schermelleh-Engel und Werner verstehen unter Reliabilität „die Genauigkeit einer Messung ... Ein Testverfahren ist perfekt reliabel, wenn die damit erhaltenen Testwerte frei von zufälligen Messfehlern sind. Das Testverfahren ist umso weniger reliabel, je größer die Einflüsse von zufälligen Messfehlern sind“ (Schermelleh & Werner 2012, S. 120). Im Zusammenhang mit der Reliabilität eines Tests stellt die Objektivität eine Voraussetzung für die Reliabilität dar. Die Reliabilität hingegen ist Voraussetzung für die Validität eines Tests (vgl. Schermelleh & Werner 2012, S. 120).

Die Reliabilität stellt die wahre Varianz der Gesamtvarianz der Testergebnisse dar. Sie bemisst also die Streuung der Merkmale in Bezug auf die wahren Testwerte. Der übrige Anteil der Gesamtvarianz kommt durch Messfehler zustande (vgl. Moosbrugger & Kelava 2012b, S. 11). Der Grad der Zuverlässigkeit (Reliabilität) eines Tests wird über einen *Reliabilitätskoeffizienten* beschrieben (vgl. Moosbrugger & Kelava 2012b, S. 11), er kann einen Wert zwischen 0 und 1 annehmen (vgl. Moosbrugger & Kelava 2012b, S. 11). Ein Wert von 1 gibt an, dass das Testergebnis ohne Messfehler zustande gekommen ist. In der Praxis würde sich ein hoher Reliabilitätskoeffizient darin widerspiegeln, dass das Testergebnis bei einer wiederholten Messung beim selben Probanden unter den gleichen Testbedingungen konstant bliebe. Ein Wert von 0 gibt genau das Gegenteil an, nämlich, dass das Testergebnis ausschließlich über Messfehler zustande gekommen ist. Für den Nachweis der Reliabilität eines Testes sollte der Wert des Reliabilitätskoeffizienten nicht unter 0,7 liegen (vgl. Moosbrugger & Kelava 2012b, S. 11). Zur Bestimmung der Reliabilität werden vier Vorgehensweisen unterschieden: (vgl. Moosbrugger & Kelava 2012b, S. 11 f.):

1. *Retest-Reliabilität*: Zur Bestimmung der Retest-Reliabilität wird dem Probanden derselbe Test zu unterschiedlichen Zeitpunkten vorgelegt. Aus den Testergebnissen der zwei verschiedenen Messzeitpunkte wird über eine Korrelation die Retest-Reliabilität angegeben (vgl. Moosbrugger & Kelava 2012b, S. 12). Bei der Interpretation des Reliabilitätskoeffizienten ist das zeitliche Intervall zwischen den Messzeitpunkten zu berücksichtigen. Es könnte durchaus sein, dass sich die Testwerte bei einer wiederholten Messung aufgrund von Erinnerungseffekten, Trainingseffekten oder sogar durch Merkmalsveränderungen unterscheiden.
2. *Paralleltest-Reliabilität*: Das Verfahren zur Bestimmung der Paralleltest-Reliabilität stellt im Gegensatz zum Verfahren zur Bestimmung der Retest-Reliabilität das genauere Verfahren dar, da es weniger anfällig gegenüber Einflussfaktoren wie Erinnerungs- und/oder Trainingseffekten oder Merkmalsveränderungen ist (vgl. Moosbrugger &

Kelava 2012b, S.12). Bei diesem Verfahren wird die Korrelation zwischen zwei Testwerten berechnet, die aus Tests mit ähnlichen Items entstehen.

3. *Testhalbierungs-Reliabilität*: Beim Verfahren der Testhalbierungs-Reliabilität wird der Test in zwei gleich große Testhälften geteilt. Zur Bestimmung der Testhalbierungs-Reliabilität wird die Korrelation der Testergebnisse aus den beiden Testhälften berechnet (vgl. Moosbrugger & Kelava 2012b, S. 12 f.). Unter Verwendung eines Korrekturfaktors wird die errechnete Korrelation der beiden Testhälften auf die Gesamtlänge des ursprünglichen Tests hochkorrigiert. Dies wird in der Regel mithilfe der Spearman-Brown-Formel durchgeführt.
4. *Innere Konsistenz*: Die Besonderheit des Verfahrens zur Bestimmung der Inneren Konsistenz ist, dass die Items als eigenständige Bestandteile des Tests gesehen werden (vgl. Moosbrugger & Kelava 2012b, S. 13). Je stärker die einzelnen Items positiv miteinander korrelieren, desto höher ist die innere Konsistenz des Tests. Die innere Konsistenz wird als Cronbach- α -Koeffizient angegeben.

In der vorliegenden Arbeit wird als Maß zur Angabe der Reliabilität der Tests auch die Andrich-Reliabilität (1982) verwendet. Die Andrich-Reliabilität wird mithilfe von Schätzungen der Personenwerte und der quadrierten Standardfehler (vgl. Eid & Schmidt 2014, S. 181 f.) über die gewichtete Maximum-Likelihood-Methode (WLE) ermittelt. Dieses Verfahren wird in dem Statistikprogramm WINMIRA angeboten. Wird ein Personenwert geschätzt, enthält dieser einen wahren Personenwert und einen Schätzfehler (vgl. Eid & Schmidt 2014, S. 181 f.). „Der Anteil der Varianz der wahren Personenwerte an der Varianz der geschätzten Personenwerte zeigt an, inwieweit die geschätzten Personenwerte wahre Personenunterschiede wiedergeben“ (Eid & Schmidt 2014, S. 181). Dieser Unterschied respektive das Verhältnis wird als Andrich-Reliabilität wiedergegeben und kann wie folgt angegeben werden:

$$(Gl. 2) \text{Rel}(\hat{\eta}) = \frac{\text{Var}(\hat{\eta}) - \text{Var}(\varepsilon)}{\text{Var}(\hat{\eta})}$$

Die Andrich-Reliabilität wird mithilfe der folgenden Formel ermittelt:

$$(Gl. 3) \quad \widehat{Rel}(\hat{\eta}) = \frac{\left[\frac{1}{n-1} \sum_{m=1}^n \left(\hat{\eta}_m - \frac{1}{n} \sum_{m=1}^n \hat{\eta}_m \right)^2 \right] - \left[\frac{1}{n} \sum_{m=1}^n \hat{\sigma}^2_{\hat{\eta}_m} \right]}{\frac{1}{n-1} \sum_{m=1}^n \left(\hat{\eta}_m - \frac{1}{n} \sum_{m=1}^n \hat{\eta}_m \right)^2}$$

Validität: Im Rahmen der Validitätsanalyse wird untersucht, ob das Instrument das Merkmal misst, das es erfassen zu wollen vorgibt (vgl. Moosbrugger & Kelava 2012b, S. 13). „Ein Test gilt dann als valide (‘gültig’), wenn er das Merkmal, das er messen soll, auch wirklich misst und nicht irgendein anderes“ (Moosbrugger & Kelava 2012b, S. 13). Die Validität stellt das wichtigste Gütekriterium dar (vgl. Moosbrugger & Kelava 2012b, S. 13). Wird eine hohe Validität eines Tests erreicht, lassen sich die Ergebnisse des Tests auf die Gesamtpopulation verallgemeinern (vgl. Moosbrugger & Kelava 2012b, S. 13 f.). Die Validität eines Tests kann in vier unterschiedlichen Bereichen getestet werden, über: die Inhaltsvalidität, die Augenscheinvalidität, die Konstruktvalidität und die Kriteriumsvalidität (vgl. Moosbrugger & Kelava 2012b, S. 15).

Inhaltsvalidität: „Unter Inhaltsvalidität versteht man, inwieweit ein Test oder ein Testitem das zu messende Merkmal repräsentativ erfasst“ (Moosbrugger & Kelava 2012b, S. 15). So sollen die Testitems eine repräsentative Auswahl aus dem Itemuniversum darstellen, um das entsprechende theoretische Konstrukt adäquat erfassen zu können (vgl. Moosbrugger & Kelava 2012b, S. 15). Die Inhaltsvalidität wird nicht über eine Kennzahl angegeben. Die inhaltliche Validität eines Items wird eher durch Expertenwissen festgelegt, das heißt, dass Experten aufgrund ihrer Fachexpertise entsprechende Items auswählen, welche das zu untersuchende theoretische Konstrukt erfassen könnten (vgl. Moosbrugger & Kelava 2012b, S. 15).

Augenscheinvalidität: Die „Augenscheinvalidität gibt an, inwieweit der Validitätsanspruch eines Tests, vom bloßen Augenschein her einem Laien ge-

rechtfertigt erscheint“ (Moosbrugger & Kelava 2012b, S. 15). Auch die Augenscheinvalidität kann nicht durch empirisch ermittelte Kennwerte angegeben werden (vgl. Moosbrugger & Kelava 2012b, S. 15 f.).

Konstruktvalidität: „Ein Test weist Konstruktvalidität auf, wenn der Rückschluss vom Verhalten der Testperson innerhalb der Testsituation auf zugrunde liegende psychologische Persönlichkeitsmerkmale («Konstrukte», «latente Variablen», «Traits») wie Fähigkeiten, Dispositionen, Charakterzüge, Einstellungen wissenschaftlich fundiert ist. Die Enge dieser Beziehung wird aufgrund von testtheoretischen Annahmen und Modellen überprüft“ (Moosbrugger & Kelava 2012b, S. 16). Es wird somit überprüft, ob mittels der verwendeten Testaufgaben auf das theoretische Konstrukt geschlossen werden kann (vgl. Moosbrugger & Kelava 2012b, S. 16).

Kriteriumsvalidität: „Ein Test weist Kriteriumsvalidität auf, wenn vom Verhalten der Testperson innerhalb der Testsituation erfolgreich auf ein «Kriterium», nämlich auf ein Verhalten außerhalb der Testsituation, geschlossen werden kann. Die Enge dieser Beziehung ist das Ausmaß an Kriteriumsvalidität (Korrelationsschluss)“ (Moosbrugger & Kelava 2012b, S. 18).

5.7 Diskussion der Methoden

Mit der klassischen Testtheorie (KTT) liegt ein geläufiger und sehr gern verwendeter Ansatz vor, um Testverfahren in ökonomischer Art und Weise konstruieren zu können. Die Praktikabilität und die relativ einfache Anwendbarkeit der Methoden innerhalb der KTT ist wahrscheinlich der primäre Grund dafür, dass sich dieser Ansatz in den angewandten Wissenschaften durchgesetzt hat und weiterhin favorisiert wird – trotz erheblicher Grenzen und Schwächen in den Bereichen der Konstruktvalidität (vgl. Moosbrugger 2012c, S. 115). Nicht nur im Bereich zur Überprüfung der Konstruktvalidität weist die KTT Schwächen auf, sondern auch bereits im konzeptionellen Bereich. Die KTT geht von der Annahme aus, dass sich der beobachtete Wert aus einem *wahren Wert* und einem *Fehlerwert* zusammensetzt. Empirisch ist es aber nicht möglich, diese beiden Größen direkt zu beobachten (vgl. Moosbrugger 2012c, S. 115). Darüber hinaus sind die Kennwerte, welche

im Rahmen der KTT Anwendung finden, stichprobenabhängig. Vor diesem Hintergrund ist es sehr problematisch, die Ausprägungen der jeweiligen Kennwerte zu verallgemeinern (vgl. Moosbrugger 2012c, S. 116). Besonders im Rahmen der Generierung von verallgemeinerbaren Itemschwierigkeiten kann dieser Umstand zu Problemen führen. Je nachdem, an welchen Probanden die Validierung der Itembatterie stattfindet, werden unterschiedliche Itemschwierigkeiten ermittelt (vgl. Moosbrugger 2012c, S. 116).

Zudem ist es nicht möglich, mithilfe von Modellannahmen die Homogenität der Items in Bezug auf das zu erfassende theoretische Konstrukt zu überprüfen (vgl. Moosbrugger 2012c, S. 115). Dies stellt aber eine zentrale Voraussetzung dar, um die Testwerte als Summenwert über alle Items angeben zu können. Ersatzweise werden die Itemtrennschärfen und die Iteminterkorrelationen bestimmt, um Aussagen über die interne Konsistenz anstellen zu können. Dies führt lediglich dazu, dass das theoretische Konstrukt in operationalisierter Weise dargestellt werden kann (vgl. Moosbrugger 2012c, S. 115).

Pospeschill (2010) gibt abschließend einen zusammenfassenden Überblick über die Vor- und Nachteile der KTT. Zu den Vorteilen der KTT zählen,

- dass sie eine gewisse Pragmatik respektive Ökonomie in Bezug auf ihre Anwendung besitzt, die auch die Zerlegung der Testwerte-Varianz erlaubt, welche die Voraussetzung für das Konzept zur Reliabilitätsprüfung darstellt (vgl. Pospeschill 2010),
- dass sie Konzepte zum Einfluss weiterer Effekte entwickelt, wie Beurteiler- oder Methodeneffekte (vgl. Pospeschill 2010).

Zu den Nachteilen der KTT zählen,

- dass unter anderem die Möglichkeit fehlt, die Axiome empirisch zu überprüfen. Wie bereits punktuell angedeutet, implizieren die Axiome eine Trennung in einen wahren Wert und in einen Fehlerwert. Darüber hinaus implizieren die Axiome eine Nullkorrelation zwischen dem wahren Wert und dem Fehlerwert sowie eine Merkmalskonstanz (vgl. Pospeschill 2010),

- dass der KTT Möglichkeiten fehlen, die Skalenniveaus der Testwerte zu überprüfen. In der KTT wird von einer Normalverteilung der Merkmale ausgegangen, welche mindestens Intervallskalenniveau besitzen müssen (vgl. Pospeschill 2010),
- dass sie Schwächen im Rahmen der Überprüfung der Konstruktvalidität besitzt. Die Bestimmung der Konstruktvalidität setzt die Homogenität der Testitems voraus und verbietet eine Stichprobenabhängigkeit (vgl. Pospeschill 2010),
- dass sie nur Rückschlüsse auf Gruppenstatistiken gewährleistet, welche Aussagen zu Wahrscheinlichkeiten auf N Elemente erlaubt. Dies verbietet, Rückschlüsse auf Wahrscheinlichkeiten in Bezug auf den Einzelfall zu treffen (vgl. Pospeschill 2010).

Daher wurde für diese Arbeit der Ansatz der Probabilistischen Testtheorie gewählt. Die aufgeführten Grenzen und Schwächen der KTT können durch die Einbeziehung der Item-Response-Theorie (IRT) überschritten respektive ausgeglichen werden. Im Unterschied zur KTT beschreibt die IRT das Reaktionsverhalten der Probanden in Abhängigkeit von den jeweiligen Personen- und Itemparametern. Zudem geht die IRT von einem Wahrscheinlichkeitszusammenhang zwischen der Merkmalsausprägung und den beobachteten Messwerten aus (vgl. Moosbrugger 2012c, S. 116).

Im Rahmen der Methodendiskussion soll auch auf das Bootstrap-Verfahren eingegangen werden. Heene et al. weisen darauf hin, dass der Einsatz des Bootstrap-Verfahrens nicht unkritisch vorgenommen werden darf. Häufig ist eine Modellannahme nach erfolgtem Bootstrap nicht korrekt (vgl. Heene et al. 2011, S. 290; Bühner 2011, S. 537). „Wird eine große Anzahl möglicher Antwortpattern für eine niedrige Itemanzahl mit dem Bootstrapverfahren generiert, dann tritt der Effekt ein, dass die Häufigkeiten der erwarteten Antwortmuster generell sehr niedrig und kaum bestimmbar sind und konsequenterweise die Power des Tests niedrig ausfällt“ (Krupp 2016, S. 211 f.).

5.8 Zusammenfassende Betrachtung des fünften Kapitels

In Kapitel 5 wurde die methodische Herangehensweise erläutert. Zur Messung des mathematischen Textverständnisses mithilfe des TeMaTex wurden insgesamt 411 Schüler aus 12 bundesweiten privaten und öffentlichen Altenpflegeschulen in die Studie eingeschlossen. Die Durchführung der Datenerhebung und Datenauswertung erfolgte nach dem Testmanual des TeMaTex. Basierend auf den Ergebnissen der Rasch-Analysen aus der ersten Erhebung, erfolgte anhand eines Expertenratings die Anpassung der Items aus dem TeMaTex an den Kontext der Altenpflege. So entstand der CareMaTex, deren Validität erneut überprüft werden musste. Hierfür wurden 253 Schüler aus fünf bundesweiten privaten und öffentlichen Altenpflegeschulen in die Untersuchung mit eingeschlossen. Die Durchführung der Datenerhebung und Datenauswertung erfolgte nach einem an den CareMaTex angepassten Testmanual. Für die Analyse der Daten wurde der probabilistische testtheoretische Ansatz präferiert, da sich die KTT unter anderem für eine Einzelfalldiagnostik wenig eignet.

6 Testtheorie und Testmodelle

In Kapitel 6.1 werden unter anderem die Grundlagen des 1-pl-Rasch-Modells beschrieben und es wird auf die Parameterschätzungen innerhalb des Modells sowie auf die relevanten Modellgeltungstests eingegangen. Abschließend erfolgt in Kapitel 6.2 eine Zusammenfassung des Kapitels.

6.1 Probabilistische Testtheorie

Das in der vorliegenden Arbeit verwendete Modell zur Bestimmung der individuellen Leistungsscores basiert auf der Item Response Theory (IRT), auch probabilistische Testtheorie (PTT) genannt (vgl. Hambleton et al. 1991; Fischer & Molenaar 1995; Rasch 1960). Ein Vorteil der IRT-Modelle respektive PTT-Modelle gegenüber den Modellen der klassischen Testtheorie (KTT-Modellen) liegt darin, dass sich die Leistungen der Personen auf einer gemeinsamen Skala abbilden lassen, obwohl diese unterschiedliche Aufgaben bearbeitet haben (vgl. Klieme et al. 2000, S. 60 f.). Zentral in der PTT ist die Unterscheidung von beobachtbaren (manifesten) und nicht beobachtbaren (latenten) Variablen. Das Lösen von Testaufgaben, egal ob richtig oder falsch, gilt als Reaktion der Testperson auf die Testaufgaben und ist somit beobachtbar. Die Testleistung einer Person stellt auch eine Beobachtung dar und ist somit – wie auch die Reaktion der Testperson – eine beobachtbare Variable. Die Fähigkeit einer Person hingegen, eine oder mehrere Testaufgaben richtig oder falsch zu lösen, ist eine latente Variable und nicht direkt beobachtbar. Diese muss über die Testleistung geschätzt werden (vgl. Klieme et al. 2000, S. 61; Knappich 2012). In der PTT werden die latenten Variablen entweder quantitativ oder qualitativ skaliert. Sind die latenten Variablen quantitativ skaliert, wird von einem Latent-Trait-(Dimension)-Modell gesprochen (vgl. Fischer & Molenaar 1995). Sind die latenten Variablen qualitativ skaliert, wird es Latent-Class-Modell genannt. Anhand der latenten Variablen werden Gruppen von Personen unterschieden (vgl. Knappich 2012).

6.1.1 Grundlagen des dichotomen 1-PL-Modells nach Rasch

Die Testleistungen der Altenpflegeschüler wurden mit dem dichotomen einparametrischen Rasch-Modell (1-PL-Modell) ausgewertet. Mit dem eindimensionalen Modell aus der Familie der Rasch-Modelle wird das mathematische Textverständnis als eine latente Dimension betrachtet. Dabei repräsentieren die einzelnen Items des TeMaTex das mathematische Textverständnis als Indikator. Das 1-PL-Modell ist ein Latent-Trait-Modell. Die Itemparameter und die Personenparameter unterscheiden sich nur von ihrer quantitativen Ausprägung voneinander. Im Rahmen des 1-PL-Modells wird versucht, die manifeste, beobachtbare Reaktion einer Testperson zu jeder Testaufgabe auf eine nicht beobachtbare, latente Personenfähigkeit und Aufgabenschwierigkeit zurückzuführen. Da die beobachtbaren Variablen dichotomer Natur sind, können diese nur die Werte zwischen 0 (nicht gelöst) und 1 (gelöst) annehmen. Auch die Werte für die Lösungswahrscheinlichkeiten $p(x = 1)$ liegen zwischen 0 und 1 (vgl. Klieme et al. 2000, S. 61 f.). Die Formel für das dichotome Rasch-Modell lautet:

$$(Gl. 4) \quad p(X_{vi} = 1) = \frac{\exp(\theta_v - \sigma_i)}{1 + \exp(\theta_v - \sigma_i)}$$

Die Formel stellt die Grundlage zur Berechnung der Lösungswahrscheinlichkeit $p(x_{vi})$ dar. $p(x_{vi})$ beschreibt die Wahrscheinlichkeit, mit der eine Aufgabe (σ) einer bestimmten Schwere mit einer vorgegebenen Personenfähigkeit (θ) gelöst werden kann. Das Indexzeichen (v) steht für die Personen und (i) für die entsprechende Testaufgabe. $\exp(x)$ steht für die Exponentialfunktion e^x , wobei e die Eulersche Zahl 2.718... darstellt. Die Beziehung zwischen der latenten Variable Personenfähigkeit (θ) und der manifesten Variable Lösungswahrscheinlichkeit $p(x_{vi})$ ist für jede Testaufgabe mithilfe der Item Characteristic Curve (ICC) grafisch darstellbar. Dabei wird in weiten Bereichen sichtbar, dass ein linearer Zusammenhang der beiden Variablen besteht. Die jeweilige Schwierigkeit einer Testaufgabe ist durch den Wendepunkt der ICC charakterisiert, bei dem die Lösungswahrscheinlichkeit $p(x_{vi}=1)$ genau 0.5 beträgt. Dies bedeutet, dass der Itemparameter und

der Personenparameter auf demselben latenten Kontinuum liegen und somit die Testaufgabe von der Testperson mit hinreichender Sicherheit gelöst werden kann (vgl. Klieme et al. 2000, S. 62; Knappich 2012). Abbildung 4 stellt dies grafisch dar.

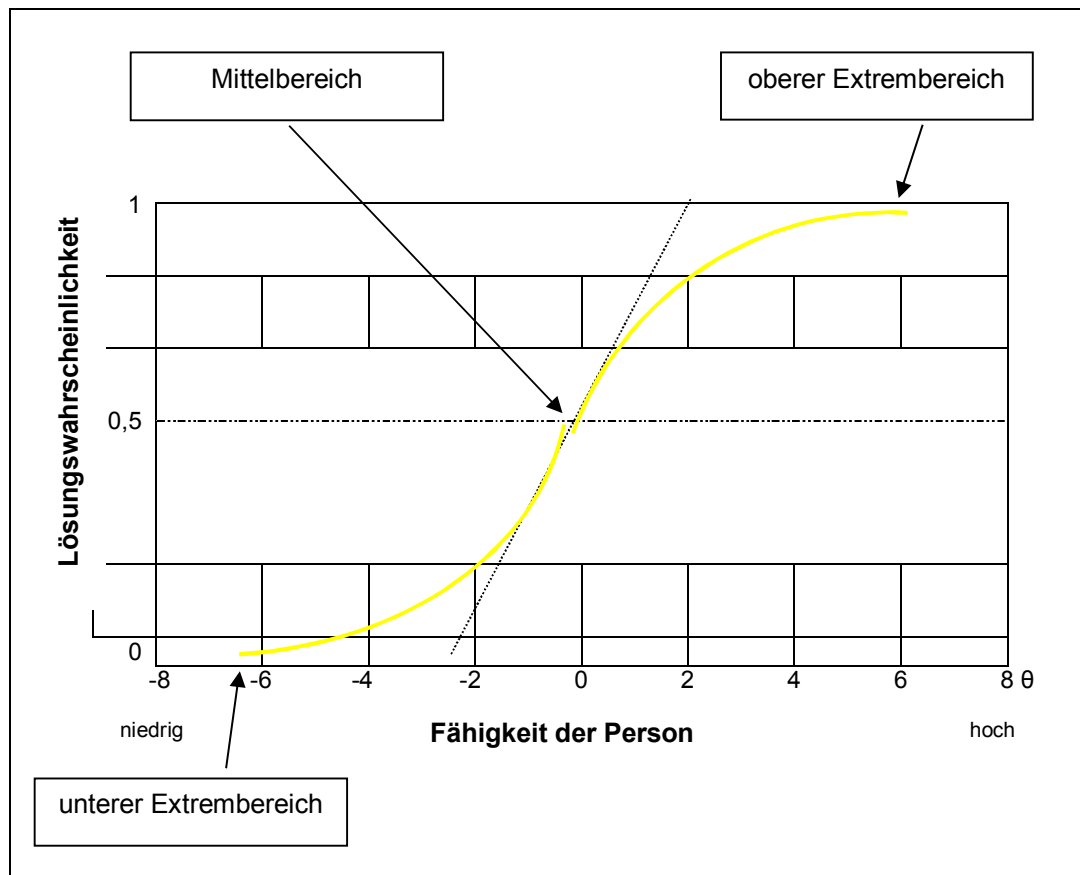


Abbildung 4: eigene Darstellung der Item Characteristic Curve 1-pl Rasch-Modell

Dem 1-PL-Rasch-Modell liegen grundsätzliche Annahmen zugrunde, sofern die empirischen Daten Rasch-konform sind: die lokale stochastische Unabhängigkeit der Testaufgaben, die Personenhomogenität und die spezifische Objektivität. Die Personenhomogenität erlaubt die Bestimmung der Aufgabenparameter unabhängig von der gewählten Stichprobe. Die spezifische Objektivität hingegen erlaubt eine Bestimmung der Personenparameter unabhängig von der Aufgabenstichprobe (vgl. Klieme et al. 2000, S. 63; Rasch 1960; Knappich 2012).

6.1.2 Parameterschätzung im Rasch-Modell

In vielen Testmodellen wird das Antwortverhalten mithilfe von Parametern (Kenngrößen) ermittelt. Parameter können die Personenfähigkeit oder die Itemschwierigkeit sein. Vielfach können die Werte dieser Kenngrößen respektive Parameter aber nicht direkt aus den beobachteten Daten berechnet werden. Oft bedarf es dafür spezieller Rechenverfahren. Im Rahmen dieser Rechenverfahren werden die Parameter anhand von Stichprobendaten geschätzt und ermittelt, wie genau dieser Schätzwert den realen Wert trifft. Das Antwortverhalten eines Probanden gibt Aufschluss über dessen Fähigkeit, die im Test enthaltenen Items zu lösen. Diese Schätzungen sind allerdings abhängig von der Größe der Itemstichprobe (Menge der Items) und der Personenstichprobe (Menge der am Test beteiligten Personen). Für die Schätzung der Personen- und Itemparameter wird das Maximum-Likelihood-Verfahren (ML-Verfahren) angewendet (vgl. Rost 2004, S. 301 ff.). Das ML-Verfahren basiert auf der Idee, „dass alle Modellparameter auf den Wert festgelegt werden, an dem die Likelihoodfunktion ihr Maximum hat“ (Rost 2004, S. 304). Dem ML-Verfahren unterliegt eine Funktion der Likelihood (Wahrscheinlichkeit). Ein weiterer Vorteil des Verfahrens liegt darin, dass der Wert der Likelihoodfunktion zudem Auskunft über die Güte des jeweiligen Testmodells gibt. Der Wert der Likelihoodfunktion ist von den Parametern des Testmodells abhängig. Sie setzt sich aus dem Produkt der Wahrscheinlichkeiten über alle Itemantworten der Personenstichprobe zusammen und bildet so die Wahrscheinlichkeit der gesamten Testdatenmatrix ab (vgl. Rost 2004., S. 301 ff.):

$$(Gl. 5) \quad L = \prod_{v=1}^N \prod_{i=1}^k p(x_{vi})$$

$$(Gl. 6) \quad P(X) = \prod_{v=1}^n \prod_{i=1}^v P(x_{vi}) = \prod_{v=1}^n \prod_{i=1}^v \left(\frac{\exp(x_{vi}(\xi - \sigma_i))}{1 + \exp(\xi - \sigma_i)} \right) = L$$

Die Funktion „beschreibt die Wahrscheinlichkeit der beobachteten Testdaten unter der Bedingung des angenommenen Testmodells als Funktion der

Modellparameter“ (Rost 2004, S. 303). Der Wert der Likelihood kann zwischen 0 und 1 variieren. Je höher der Wert der Likelihood ist, desto wahrscheinlicher ist es, die richtigen Schätzer für die Parameter des Testmodells bestimmt zu haben (vgl. Moosbrugger 2012d, S. 239 ff.).

Schätzung der Itemparameter: In dieser Arbeit finden verschiedene Methoden zur Schätzung von Itemparametern Anwendung, welche im Folgenden näher erläutert werden.

cML-Methode: Das in dieser Arbeit verwendete Verfahren zur Schätzung der Itemparameter findet unter anderem in den Statistik-Programmen WINMIRA und eRm Anwendung. Die CML-Schätzung basiert auf der Maximum-Likelihood-Methode. Aufgrund der Annahme, dass es sich bei den Summenwerten der Personen um erschöpfende Statistiken der Personenfähigkeit handelt, ist es möglich, den Personenparameter aus der Gleichung zu streichen (vgl. Bühner 2011, S. 500). „Sie wird deshalb so genannt, weil Wahrscheinlichkeiten auf die Zeilenrandsumme bzw. die Summenwerte zurückgeführt bzw. konditioniert werden ... Die konditionale bzw. bedingte Likelihoodfunktion (cL) kann nur im Rasch-Modell für die Schätzung der Itemparameter herangezogen werden, da eine solche Konditionierung nur mithilfe von Summenwerten sinnvoll ist, die erschöpfende Statistiken der Personenfähigkeit darstellen“ (Bühner 2011, S. 500).

Pair-Wise Algorithmus: Dieser Schätzalgorithmus wird im Rahmen des Statistik-Programms RUMM 2030 verwendet und beruht nicht auf dem Prinzip der Maximum-Likelihood-Schätzung. Er betrachtet hingegen Itempaare, nach denen auch das Verfahren bezeichnet wurde. Die Schätzwerte, die durch diesen Algorithmus gebildet werden, unterscheiden sich nur geringfügig von den Schätzwerten, welche die CML-Methode generiert (vgl. Rost 2004, S. 310 f.).

Schätzung der Personenparameter: In dieser Arbeit findet eine spezielle Methode zur Schätzung der Personenparameter Anwendung, welche im Folgenden näher erläutert wird.

WLE-Methode: Für die Schätzung der Personenparameter gibt es unterschiedliche Verfahren, die verschiedene Ergebnisse produzieren, was eine Vergleichbarkeit erschwert (vgl. Rost 2004, S. 309). Zur Schätzung der Personenparameter wird in dieser Arbeit die Weighted-Likelihood-Estimates (WLE) verwendet. Die WLE eignet sich gut, um für extreme Summenwerte Schätzungen für Personenparameter durchzuführen (vgl. Bühner 2011, S. 501). Mithilfe eines Korrekturterms, welcher in die MLE-Formel integriert wird, ist es möglich, „endliche Werte für die Personenparameter der Extremscores zu schätzen. Der Korrekturterm führt schließlich dazu, dass betragsmäßig kleine Personenparameter (nahe Null) innerhalb der WLE-Schätzung gegenüber der MLE-Schätzung vergrößert werden – betragsmäßig große Personenparameter werden hingegen verkleinert“ (Jordan 2011a, S. 143; vgl. Rost 2004, S. 314).

6.1.3 Parameterschätzung bei missing data

Im Rahmen der Datenerhebung kann es vorkommen, dass nicht alle Personen alle Items des TeMaTex beantworten. Es liegen also fehlende Werte (missing data) in der Datenmatrix vor. Fehlende Werte können in zwei Kategorien eingeteilt werden: zufällig auftretende fehlende Werte und nicht zufällig auftretende Werte. Die zufälligen missing data können wiederum in designbedingte missing data oder missing-at-random (MAR) unterschieden werden. In solchen Fällen gibt es eine spezielle und einfache Methode, um die Lücken im Datensatz adäquat zu schließen. Diese Methode wird Imputationstechnik genannt, bei der die bestehenden Lücken mit Pseudodaten gefüllt werden (vgl. Rost 2004, S. 324 f.).

Im vorliegenden Fall handelt es sich jedoch um nicht zufällige missing data, weil sich die fehlenden Daten nicht auf das Design begründen, sondern die Ursache bei der Testperson selbst liegt. Dies bedarf weiterer Methoden zur Behandlung von fehlenden Werten. Bei Rasch-Modellen wie dem dichotomen 1-PL-Rasch-Modell stellt der Umgang mit fehlenden Werten im Rahmen der Itemparameterschätzung keine größere Herausforderung dar.

Durch einen paarweisen Algorithmus werden nur die Personen berücksichtigt, die bei einem Itempaar beide Items bearbeitet haben. Sind die Itemparameter einmal geschätzt, ist die Schätzung der Personenparameter problemlos. Mit folgender Gleichung werden nur die bearbeiteten Items berücksichtigt:

$$(Gl. 7) \quad r_v = \sum_{i=1}^{k_v} P_{vi}$$

In der Formel steht k_v für die Anzahl der Items, v für die Person, die die Items vollständig bearbeitet hat und i ist der Index des bearbeiteten Items (vgl. Rost 2004, S. 325 f.).

6.1.4 Differenz zwischen Personen- und Itemparameter

In Kapitel 12.1 und 12.2 werden Kompetenzstufeneinteilungen vorgenommen, um die Leistungen der Schüler in eine entsprechende Kompetenzstufe einordnen zu können. Um genauere Angaben machen zu können, ab wann ein Proband eine nächsthöhere Kompetenzstufe erlangen kann, wurde eine 70-prozentige Wahrscheinlichkeit gewählt. Diese Vorgehensweise wird auch bei großen Studien wie der PISA-Studie gewählt (vgl. Jordan 2011a, S. 352). Dies bedeutet, dass aus jedem Personenparameter die entsprechenden Itemparameter errechnet werden müssen. Die dafür verwendete Formel wird aus der Modellgleichung des Rasch-Modells berechnet. Die Modellgleichung des Rasch-Modells sowie die daraus abgeleitete Formel zur Berechnung der Differenz werden im Folgenden dargestellt:

$$(Gl. 8) \quad p(x_{vi} = 1) = \frac{\exp(x_{vi}(\theta_{vi} - \sigma_i))}{1 + \exp(\theta_{vi} - \sigma_i)}$$

$$(Gl. 9) \quad d = \ln \frac{p}{1 - p}$$

Der Buchstabe d steht dabei für die logarithmierte Differenz zwischen der Zahl 1 und der gewählten Wahrscheinlichkeit $p = 0,7$. Diese Differenz be-

trägt 0,847. Somit muss der Itemparameter immer kleiner sein als der vorgegebene Personenparameter (vgl. Jordan 2011a, S. 352). Die ausführlichen Rechenschritte zur Berechnung der Differenz zwischen der Personen- und Itemschwierigkeit können Anhang 5 entnommen werden.

6.1.5 Modellgeltungstests bei dichotomem 1-PL-Rasch-Modell

Modellgeltungstests werden verwendet, um darzustellen, inwieweit das für die Datenauswertung ausgewählte Modell zu den erhobenen Daten passt. Bei den meisten Modellgeltungstests wird vorausgesetzt, dass bereits die Item- und Personenparameter des Tests vorliegen. Im Allgemeinen lässt sich feststellen, dass fast alle der probabilistischen Testmodelle mehr oder weniger gut zu den Daten passen. Es gestaltet sich schwierig, Kriterien heranzuziehen, ab wann ein Modellgeltungstest die Daten nicht ausreichend erklärt respektive beschreibt. Oftmals geschieht dies willkürlich (vgl. Rost 2004, S. 330). Es ist möglich, mit einem Modelltest verschiedene Modellannahmen gleichzeitig zu überprüfen. Zu diesen Annahmen gehören die Eindimensionalität des zu untersuchenden Konstruktes, die lokale stochastische Unabhängigkeit der Itemreaktionen, die Homogenität der Items, die Suffizienz der Summenscores und die konstanten Trennschärfen der Items. Mithilfe des Andersen-Tests und des Wald-Tests kann die Personenhomogenität beziehungsweise die Itemhomogenität überprüft werden (vgl. Rost 2004, S. 345).

6.1.5.1 Grafischer Modellgeltungstest

Im Rahmen der Prüfung auf Personenhomogenität wird die Personenstichprobe zweigeteilt. Teilungskriterium ist der Summenscore der Personen beziehungsweise die Schätzwerte der Itemparameter. In Form eines grafischen Modelltests lassen sich die Werte beider Gruppen miteinander vergleichen. Im Idealfall bilden die Punkte in der Graphik eine Gerade im 45°-Winkel (vgl. Rost 2004, S. 347 f.). Je weiter die Punkte von der Geraden entfernt liegen, „desto ausgeprägter ist die Unterschiedlichkeit der beiden Personengruppen hinsichtlich der relativen Schwierigkeiten der Items“ (Rost 2004, S. 348).

Der grafische Modelltest ersetzt keinen statistischen Modellgeltungstest. Er gibt lediglich eine Richtung vor, indem er aufzeigt, welche Items relativ leicht und relativ schwer von den Personengruppen beantwortet wurden. Einen statistischen Modellgeltungstest stellt der bedingte Likelihood-Quotienten-Test von Andersen dar (vgl. Rost 2004, S. 348).

6.1.5.2 Wald-Test

Zur Überprüfung der Itemhomogenität wird in der vorliegenden Untersuchung der Wald-Test verwendet. Die Grundidee des Tests liegt darin, dass sich bei Geltung des Rasch-Modells der geschätzte Aufgaben-Parameter eines einzelnen Items nicht signifikant zwischen den aus der Gesamtstichprobe gewonnenen Teilstichproben unterscheiden darf. Die Teilung der Gesamtstichprobe wird meist anhand des Mittelwertes, Median, Alters oder Geschlechts aus den Rohwerten der Gesamtstichprobe vorgenommen. Wird ein Item signifikant, müsste es aus dem Test entfernt oder durch eine andere Aufgabe ersetzt werden. Der Wald-Test ähnelt dem Likelihood-Quotienten-Test (vgl. Strobel 2010, S. 44 f.).

6.1.5.3 Tests zur Überprüfung der Signifikanz

Der Signifikanztest prüft die Nullhypothese (H_0) – dass das Rasch-Modell gilt – gegen die Alternativhypothese (H_1) – dass das Rasch-Modell nicht gilt. Ist die Testung des Modells signifikant ($p < 0.05$), muss das getestete Modell abgelehnt werden, also H_0 verworfen und H_1 angenommen werden kann (vgl. Rost 2004, S. 331). Der Likelihood-Quotienten-Test nach Andersen ist für diese Untersuchung keine aussagekräftige Messmethode, da die nötige Voraussetzung nicht gegeben ist. Hierfür müssten alle möglichen Antwortmuster vorliegen, was aufgrund des zu geringen Stichprobenumfangs nicht möglich ist. Vor diesem Hintergrund muss auf das Bootstrap-Verfahren zurückgegriffen werden, um eine fiktive Modellgeltung vorhersagen zu können (vgl. Rost 2004, S. 331 ff.).

6.1.5.3.1 Likelihood-Quotienten-Test

Im Rahmen der Modellprüfung nach dem Likelihood-Quotienten-Test liegen zwei unterschiedliche Varianten vor. Für beide Varianten wird eine χ^2 -Prüfgröße errechnet. Der Anderson-Test gibt vor, dass die Stichprobe in zwei Teilstichproben aufgeteilt werden muss. Für jede Teilstichprobe werden die Likelihoods errechnet. Anschließend kann, basierend auf den beiden Likelihoods, die χ^2 -Prüfgröße ermittelt werden (vgl. Rost 2004, S. 334). Bei der zweiten Variante wird der Likelihood-Quotienten-Test gegen das saturierte Modell getestet. Das saturierte Modell ist das Modell, das die Daten perfekt beschreibt. Auch hier wird eine χ^2 -Prüfgröße ermittelt, in dem die Likelihood des saturierten Modells berechnet wird (vgl. Rost 2004, S. 334).

6.1.5.3.2 Pearson χ^2 Test

Mithilfe des Pearson- χ^2 -Tests werden die erwarteten und beobachteten Häufigkeiten der Antwortmuster überprüft. Liegen große Abweichungen zwischen den beobachteten und erwarteten Häufigkeiten der Antwortmuster vor, ist von keiner Modellgeltung auszugehen. Angegeben wird die Teststatistik mit einer χ^2 -Prüfgröße in Form eines p -Wertes (vgl. Bühner 2006, S. 346).

6.1.5.3.3 Bootstrap-Verfahren

Mithilfe des Bootstrap-Verfahren wird eine künstliche Datenmatrix simuliert. So können die beobachteten Werte auf die Geltung des Rasch-Modells überprüft werden, obwohl die Voraussetzungen zur Ermittlung der χ^2 -Prüfgröße für die vorliegende Stichprobe nicht gegeben sind. Hierfür müssen zunächst die Item- und Personenparameter sowie die χ^2 -Prüfgröße der originären Stichprobe ermittelt werden. Anschließend werden viele Datensätze mit demselben Stichprobenumfang wie die der originären Stichprobe simuliert und die dazugehörigen χ^2 -Prüfgrößen ermittelt. Die Statistiksoftware WINMIRA kann bis zu 400 Datensätze simulieren. Danach werden die fiktiv ermittelten χ^2 -Prüfgrößen mit der χ^2 -Prüfgröße der originären Stich-

probe verglichen. Es wird verglichen, wie viele der simulierten χ^2 -Prüfgrößen kleiner sind als die χ^2 -Prüfgröße der originären Stichprobe (vgl. Bühner 2006, S. 347; Rost 2004, S. 337). „Gehört der empirisch beobachtete χ^2 -Wert zu den 5 Prozent höchsten χ^2 -Werten der simulierten Verteilung, wird das Modell abgelehnt. Es ist unter diesen Gegebenheiten unwahrscheinlich, dass der χ^2 -Wert, der empirisch beobachtet wurde, zu den χ^2 -Werten gehört, auf die das Modell passt“ (Bühner 2006, S. 347). Die Testsignifikanz wird bei WINMIRA mit p angegeben (vgl. Bühner 2006, S. 348). Nach dem gleichen Prinzip lässt sich auch die χ^2 -Prüfgröße nach der Cressie-Read-Statistik berechnen. Mit der Pearson-Chi²-Statistik und der Cressie-Read-Statistik wird überprüft, ob die Häufigkeitsverteilungen der beobachteten Antwortmuster mit den theoretisch erwarteten Antwortmustern übereinstimmen, und zwar unter der Annahme, dass das Rasch-Modell gilt (vgl. Heene et al. 2011, S. 286).

6.1.5.3.4 Lokale Modellgeltungstests

Personen-Fit-Index: Die Ursache einer nicht vorhandenen Modellgeltung können auch nicht modellkonforme Antworten der Probanden sein. Bemerkbar macht sich dies z. B. in dem widersprüchlichen Antwortverhalten von Probanden, die nicht die leichten Items, dafür aber die schwersten Items lösen können (vgl. Moosbrugger 2012d, S. 246). Liegen die Personen-Fit-Indizes im kritischen Bereich, sollten diese Items aus dem Test entfernt werden (vgl. Bühner 2006, S. 367). „Der Index beruht auf der bedingten Wahrscheinlichkeit des Antwortmusters einer Person, wobei die Bedingung ihr Summenscore r_v ist“ (Rost 2004, S. 363). Der Personen-Fit-Index gibt an, mit welcher Wahrscheinlichkeit eine Person mit einem entsprechendem Summenscore genau jenes Antwortmuster erreicht (vgl. Rost 2004, S. 363). Folgende Formel soll dies verdeutlichen (vgl. Rost 2004, S. 364):

$$(Gl. 10) \quad p(\underline{x}|\underline{r}) = \frac{\exp\left(-\sum_{i=1}^k X_i \sigma_i\right)}{\sum_{\underline{x}|\underline{r}} \exp\left(-\sum_{i=1}^k X_i \sigma_i\right)}$$

Zunächst wird die Verteilung der bedingten Patternwahrscheinlichkeiten für jeden Score r ermittelt. Anschließend werden die Häufigkeiten der beobachteten Antwortmuster berechnet. Die Wahrscheinlichkeit eines Antwortpattern wird in Form eines z-Wertes angegeben (vgl. Rost 2004, S. 364).

Item-Fit-Index: „Itemfit-Maße zeigen an, wie gut ein Item zu dem Testmodell passt, das als Grundlage der Testauswertung dient“ (Rost 2004, S. 371). Items mit niedriger Trennschärfe bekommen einen schlechten Item-Fit, Items mit einer hohen Trennschärfe erhalten einen guten Item-Fit (vgl. Rost 2004, S. 371). Grundsätzlich lassen sich zwei unterschiedliche Ansätze diskutieren: der residuen-orientierte und der likelihood-orientierte Ansatz (vgl. Rost 2004, S. 371).

Residuen-orientierter Ansatz: „Als Residuen bezeichnet man die Abweichungen von beobachteten und theoretisch erwarteten Werten. Gemeint sind im vorliegenden Fall die beobachteten und erwarteten Itemantworten, x_{vi} und $Erw(x_{vi})$ “ (Rost 2004, S. 371). Bei diesem Ansatz wird also die Differenz der beobachteten und der erwarteten Itemantworten durch die Standardabweichung der erwarteten Itemantworten geteilt, um standardisierte Abweichungen zu erhalten (vgl. Rost 2004, S. 371). Folgende Formeln sollen dies verdeutlichen (vgl. Rost 2004, S. 372).

$$(Gl. 11) \quad z_{vi} = \frac{(x_{vi} - Erw(x_{vi}))}{\sqrt{Var(x_{vi})}} \quad (Gl. 12) \quad z(Q_i) = \frac{Q_i - Erw(Q_i)}{\sqrt{Var(Q_i)}}$$

Die Berechnungen für die standardisierten Residuen können für jede einzelne Itemantwort vollzogen werden (vgl. Rost 2004, S. 372). Hierbei handelt es sich um z-Werte, also „um standardnormalverteilte Größen mit Erwartungswert 0 und Standardabweichung 1“ (Rost 2004, S. 372).

Likelihood-orientierter Ansatz: „Likelihood-basierte Fit-Maße gehen dagegen von der Wahrscheinlichkeit beobachteter Daten aus“ (Rost 2004, S. 371). Im Interesse stehen im Speziellen die Wahrscheinlichkeiten der beobachteten Spaltenvektoren der jeweiligen Items (Q_i). Es wird überprüft, wie viele Personen das jeweilige Item lösen konnten und ob das jeweilige Item

von den richtigen Personen gelöst wurde (vgl. Rost 2004, S. 373). Die Berechnungen führen auch zu z-Werten. Die Residuen werden durch die Wahrscheinlichkeiten der Spaltenvektoren ersetzt (vgl. Rost 2004, S. 373). „Bei den ZQ-Werten handelt es sich um eine z-verteilte Prüfgröße. Mit ihr kann bestimmt werden, ob bei den einzelnen Items eine Abweichung vom erwarteten Lösungsmuster vorliegt oder nicht. Diese Abweichung liegt dann vor, wenn für ZQ-Werte extreme Werte – auf 5%-Niveau betragsmäßig größer als 1,96 – ermittelt werden. Negative ZQ-Werte stehen dabei für einen Item-Underfit, während positive Werte für ZQ einen Item-Overfit signalisieren. [...] WINMIRA kennzeichnet die Übertrittswahrscheinlichkeiten auf 1%-Niveau (!) und auf 5%-Niveau (?) mit der jeweiligen Angabe, ob ein Item-Underfit (-) oder ein Item-Overfit (+) vorliegt“ (Jordan 2011a, S. 246).

6.1.6 Differential Item Functioning

“It is necessary to check whether the items work in the same way for different groups of people who might be compared” (Andrich 2009, S. 31). Um dieses Postulat von Andrich zu erfüllen, gibt es verschiedene Möglichkeiten. Im Kern aller Verfahren zur Feststellung eines *Differential Item Functioning* (DIF) geht es darum, ob sich die geschätzten Aufgaben-Parameter einer Personen-Stichprobe in verschiedenen Gruppen, wie zum Beispiel zwischen Haupt- und Realschülern, systematisch unterscheiden. Bei Geltung des Rasch-Modells dürfen sich die Aufgaben-Parameter der jeweiligen Gruppen nicht systematisch voneinander unterscheiden (vgl. Strobel 2010, S. 39).

6.2 Zusammenfassende Betrachtung des sechsten Kapitels

In Kapitel 6 wurde der gewählte probabilistische testtheoretische Ansatz erläutert. Aufgrund der dichotomen Daten wurde das 1-PL-Rasch-Modell gewählt. Basierend auf der dichotomen Auswertungsstruktur (0 = nicht gelöst/nicht bearbeitet; 1 = richtig gelöst) lagen keine fehlenden Daten vor. Die Schätzung der Parameter in Bezug auf die Itemschwierigkeit wurde je nach Statistik-Software entweder über die CML-Schätzung oder über den Pairwise-Algorithmus durchgeführt. Die Schätzung der Parameter bezüglich der

Personenfähigkeit erfolgt unter anderem über die WLE. Die Überprüfung der Personenhomogenität erfolgte über den Andersen-Test sowie über den grafischen Modelltest. Der Andersen-Test, der Pearson- χ^2 -Test in Verbindung mit dem Bootstrap-Verfahren dienten zudem zur Prüfung der Modellpassung. Die Itemhomogenität wurde mithilfe des Wald-Tests überprüft.

7 Test zur Erfassung des mathematischen Textverständnisses (TeMaTex)

Kapitel 7.1 enthält eine allgemeine Beschreibung des TeMaTex. Hierbei wird der Aufbau des Tests dargestellt. In Kapitel 7.2 erfolgt eine Erläuterung zu den Gütekriterien Objektivität, Reliabilität und Validität des Tests. Kapitel 7.3 stellt eine Zusammenfassung des Tests vor allem in Bezug auf wichtige statistische Kennzahlen dar.

7.1 Beschreibung des Tests und der Aufgaben

Der *Test zur Erfassung des mathematischen Textverständnisses* (TeMaTex) misst das mathematische Textverständnis von Schülern, die sich kurz vor, während oder kurz nach dem Übergang vom allgemeinbildenden zum berufsbildenden Schulsystem befinden (vgl. Jordan & Stein 2011b). Der TeMaTex ist als Einzel- oder Gruppentest anwendbar und kann in den Klassenstufen 9 und 10 sowie in allen berufsbildenden Schulen als Diagnoseinstrument eingesetzt werden. Getestet wird mit jeweils sieben verschiedenen Aufgabentexten. Drei davon bestehen aus einer Kombination von kontinuierlichen Texten wie Fließtexten und nicht kontinuierlichen Texten wie Tabellen und Abbildungen. Die restlichen Aufgabentexte sind kontinuierliche Texte. Insgesamt beinhaltet der TeMaTex 18 Aufgaben (vgl. Jordan & Stein 2011b). Im Folgenden wird die Einteilung der Aufgaben in die unterschiedlichen Kompetenzstufen beschrieben.

Einteilung der Aufgaben in unterschiedliche Kompetenzstufen: Die Aufgaben aus dem Test zur Erfassung des mathematischen Textverständnisses können unterschiedlichen Kompetenzstufen zugeordnet werden (vgl. Jordan 2011a, S. 350 f.). Abbildung 5 stellt dies grafisch dar und zeigt drei statistisch generierte Kompetenzstufen auf, welche zur Klassifizierung des latenten Merkmals *mathematisches Textverständnis* dienen sollen. Den einzelnen Kompetenzstufen wurden die 18 Aufgaben aus dem TeMaTex aufgrund ihrer Itemschwierigkeit zugewiesen.

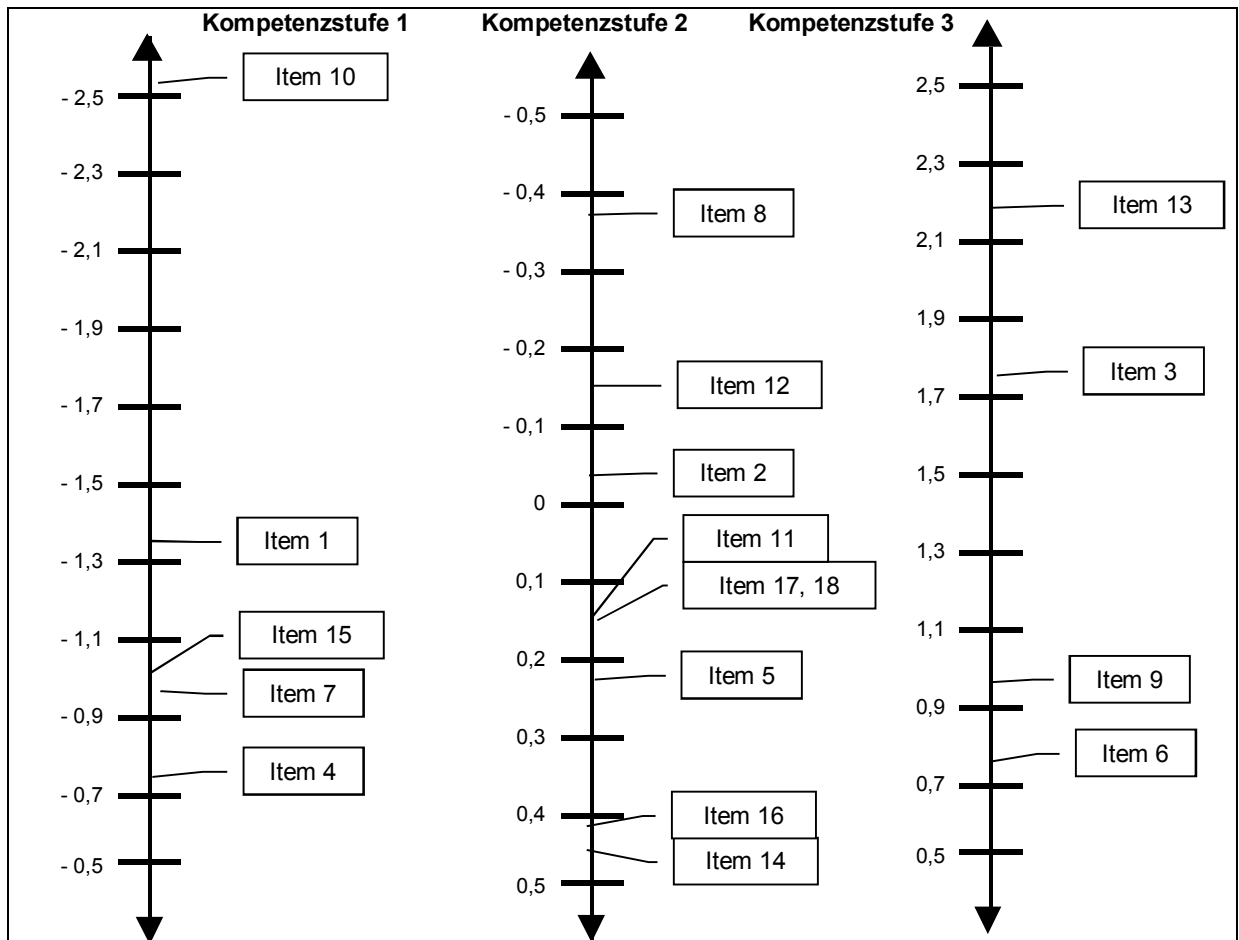


Abbildung 5: Statistische Darstellung der Kompetenzstufen des TeMaTex (nach Jordan 2011a, S. 350)

Die Kompetenzstufe 1 stellt die Stufe dar, in der die einfachsten Aufgaben zu finden sind. In statistischen Kennwerten ausgedrückt, werden insgesamt fünf Items des TeMaTex mit einer Itemschwierigkeit σ_i von weniger als -0,5 der Stufe zugeordnet. Die Kompetenzstufe 2 umfasst neun Items mit einer Schwierigkeit zwischen -0,5 bis 0,5. Die dritte Kompetenzstufe ist das Level, in dem Items des TeMaTex mit einer Schwierigkeit von mehr als +0,5 eingruppiert werden. Sie umfasst vier Items (vgl. Jordan 2011a, S. 350). Um den einzelnen Kompetenzstufen inhaltlich eine Beschreibung zuordnen zu können, wurden die Items innerhalb einer Stufe miteinander verglichen und als Ergebnis die Gemeinsamkeiten in Form einer Beschreibung der Kompetenzstufe generiert (vgl. Jordan 2011a, S. 351). In der Tabelle 14 werden die einzelnen Kompetenzstufen inhaltlich beschrieben.

Tabelle 14: Inhaltliche Beschreibung der Kompetenzstufen (nach Jordan 2011a, S. 351)

Beschreibung der Kompetenzstufen	
Stufe 1 $(- 2,5 \leq \sigma_i \leq - 0,5)$	<p>Bei Aufgaben der ersten Kompetenzstufe wird von den Probanden verlangt, dass sie aus dem Textmaterial zwei natürliche Zahlen identifizieren. Anschließend muss eine Addition oder Multiplikation durchgeführt werden. Die Zahlen sind im Material leicht zu finden und die Rechenoperationen sind ebenfalls einfach. Im mathematischen Lösungsprozess muss nur eine einzige Rechenoperation durchgeführt werden.</p>
Stufe 2 $(- 0,5 \leq \sigma_i \leq + 0,5)$	<p>Bei Aufgaben der zweiten Kompetenzstufe müssen aus dem Textmaterial zwei oder mehr Zahlen identifiziert werden. Aufgrund vieler konkurrierender Informationen sind die benötigten Zahlen jedoch nicht direkt ersichtlich, sodass zum Auffinden der benötigten Informationen ein umfassendes Textverständnis notwendig ist. Die mathematischen Operationen sind zudem schwerer als bei Aufgaben der ersten Kompetenzstufe. So werden beispielsweise Kenntnisse der Bruchrechnung, Prozentrechnung und Dezimalbruchrechnung verlangt, während bei Aufgaben der ersten Kompetenzstufe ausschließlich mit natürlichen Zahlen gerechnet wird. Meist muss im Lösungsprozess ein einziger Rechenschritt durchgeführt werden. Es gibt jedoch auch Aufgaben, bei denen mehr als ein Rechenschritt durchzuführen ist.</p>
Stufe 3 $(+ 2,5 \geq \sigma_i \geq + 0,5)$	<p>Aufgaben der dritten Stufe zeichnen sich dadurch aus, dass sämtliche Textmaterialien (Fließtext, Diagramm, Tabelle) zur Bearbeitung der Aufgaben hinzugezogen werden müssen. Mehrere Informationen müssen dabei miteinander verknüpft werden. Die Suche nach den benötigten Informationen wird dadurch erschwert, dass in den Texten eine oder mehrere konkurrierende Informationen gegeben sind. Für den Lösungsprozess sind in der Regel mehr als zwei Rechenschritte notwendig.</p>

Im Folgenden werden exemplarisch ein Aufgabentext sowie eine dazugehörige Aufgabe aus dem TeMaTex dargestellt und inhaltlich beschrieben. Alle weiteren Aufgabentexte und Aufgaben können dem TeMaTex (Jordan 2011c) entnommen werden. Im ersten Aufgabentext wird mit einem Fließtext die Ausgangssituation beschrieben. Der zweite Fließtext dient der Beschreibung eines Ausschnitts aus einem Vertrag zwischen der Stadt Karls-

ruhe und der Firma RenoAG zur Aufschlüsselung der Kosten für die Renovierungsarbeiten. Darüber hinaus enthält der Aufgabentext einen Raum- und Inventarplan (vgl. Jordan 2011a, S. 199).

Text 1: Die Renovierungsfirma

Für die Renovierung und Instandhaltung der Räumlichkeiten der Stadt Karlsruhe ist die Firma RenoAG zuständig. Im Jahr 2009 standen einige Renovierungsarbeiten an. Die Stadt Karlsruhe hat mit der Firma RenoAG einen Vertrag abgeschlossen, der die Kosten bzgl. der Renovierungsarbeiten regelt (vgl. Vertragsauszug unten).

Bearbeiten Sie die Aufgaben auf der Rückseite zu diesem Text. Beachten Sie dazu den folgenden Vertragsauszug und die Tabelle auf der folgenden Seite (rechts).

Vertragsauszug zwischen der Firma RenoAG und der Stadt Karlsruhe:

§ 5: Kosten für Renovierungsarbeiten:

(1) Die Bezahlung für durchgeführte Renovierungsarbeiten richtet sich nicht nach der Arbeitszeit, sondern wird anhand des Materialverbrauchs berechnet. Demnach stellt die Firma RenoAG der Stadt Karlsruhe folgende Beträge in Rechnung:

- pro m² verlegten Teppichs: 15 €
- pro m verlegter Bodenleisten: 6,50 €
- pro m² gestrichener Wand: 1,30 €.

In diesen Beträgen sind die Arbeitskosten bereits enthalten.

(2) Die Firma RenoAG erstellt die Rechnungen für Renovierungsarbeiten auf Grundlage der Tabelle „Raum- und Inventarplan der Stadt Karlsruhe“. Bei den zu streichenden Wandflächen werden Türen und Fensterflächen nicht einzeln berechnet, sondern es wird pauschal 20 % der gesamten Wandfläche abgezogen. Die Decke wird nicht gestrichen.

Abbildung 6: Aufgabentext 1 des TeMaTex (nach Jordan & Stein 2011c, S. 2)

Tabelle 15: Raum- und Inventarplan der Stadt Karlsruhe (Erdgeschoss und 1. Obergeschoss) (nach Jordan & Stein 2011c, S. 3)

	Erdgeschoss										1. Obergeschoss									
	Raum 0.01	Raum 0.02	Raum 0.03	Raum 0.04	Raum 0.05	Raum 0.06	Raum 0.07	Raum 0.08	Raum 0.09	Raum 0.10	Raum 1.01	Raum 1.02	Raum 1.03	Raum 1.04	Raum 1.05	Raum 1.06	Raum 1.07	Raum 1.08	Raum 1.09	Raum 1.10
Angaben zum Raum (alle Räume eine rechteckige Grundfläche)																				
Länge (in m)	8	6,5	7	4,7	12	15	6	8	5	4	7	4	5	20	5	3	6	12	5	7
Breite (in m)	6	4	5,5	5	12	20	7	4	3	6	5	9	6	15	5	6	4	17	6	5
Höhe (in m)	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
Anzahl der Türen	1	1	2	2	1	2	1	2	1	1	1	1	1	2	1	2	1	3	1	1
Breite pro Tür (in m)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Angaben zum Inventar im Raum																				
Anzahl Regale	4	3	3	5	4	7	4	2	2	3	4	3	4	8	3	2	1	4	2	2
Anzahl Schränke	2	1	2	2	4	5	3	2	2	1	2	3	1	4	2	3	1	3	2	1
Anzahl Stühle	5	5	7	2	7	16	6	4	3	2	4	6	8	15	3	6	5	16	6	8
Anzahl Tische	2	2	4	2	7	12	3	3	2	2	4	3	5	14	3	4	3	11	3	5
Anzahl Computer	3	2	5	2	10	14	5	2	1	0	3	4	3	10	5	3	2	8	6	5
Anzahl Monitore	3	2	10	4	16	14	5	3	1	0	4	5	3	10	6	3	2	10	6	7

1. Aufgabe:

Der Raum 0.06 wird komplett neu renoviert. Berechnen Sie, wie viel m² Teppich verlegt werden müssen.

Abbildung 7: Item 1 zum Aufgabentext 1 (nach Jordan & Stein 2011c, S. 4)

Bei der Aufgabe 1 muss eine rechteckige Grundfläche bestimmt werden, um einen Teppich verlegen zu können. Hierfür müssen die Probanden die Maße des Raumes aus der Tabelle entnehmen. Die Rechenoperation zur Bestimmung einer rechteckigen Grundfläche ist nicht direkt aus der Aufgabe ersichtlich (vgl. Jordan 2011a, S. 199 f.).

2. Aufgabe:

Berechnen Sie, wie viel m Bodenleisten im Raum 0.06 verlegt werden müssen. Beachten Sie dabei, dass die Breite der Türen abgezogen werden muss (vgl. Tabelle)!

Abbildung 8: Item 2 zum Aufgabentext 1 (nach Jordan & Stein 2011c, S. 4)

Bei Aufgabe 2 muss berechnet werden, wie viel Meter Fußleisten zu verlegen sind. Hierfür muss vom Proband zunächst der Umfang des Raumes bestimmt werden. Anschließend müssen von dem Ergebnis die Türbreiten abgezogen werden. Die Aufgabe bedarf mehrerer Rechenoperationen. Der Tabelle müssen weitere Informationen entnommen werden. Aufgabe 2 ist schwieriger als Aufgabe 1 und kann vom Probanden nur durch Schlussfolgerungen gelöst werden, weil die Rechenoperation nicht aus der Aufgabe ersichtlich ist (vgl. Jordan 2011a, S. 200). Im Folgenden werden die Ergebnisse zu den Aufgabenanalysen der verbleibenden 16 Items des TeMaTex zusammengefasst dargestellt.

Bei Aufgabe 3 muss der Proband berechnen, wie groß die zu streichende Wandfläche ist. Hierfür muss die Mantelfläche des Raumes errechnet werden. Fenster und Türen dürfen nicht eingerechnet werden. Auch bei Aufgabe 3 sind mehrere Rechenoperationen notwendig, um die Aufgabe lösen zu können (vgl. Jordan 2011a, S. 200). Möchte ein Proband die Aufgaben 4 bis 6 richtig lösen, muss er anhand der Flächen- und Längenangaben die Preise für die Renovierung berechnen. Hierfür muss er Informationen aus der Tabelle mit den entsprechenden Informationen aus dem Vertragsaus-

zug kombinieren. Die Rechenoperationen der Aufgaben sind teilweise etwas komplizierter, da ganze Zahlen mit Kommazahlen multipliziert werden müssen (vgl. Jordan 2011a, S. 200). Tabelle 16 verdeutlicht die Eingruppierung der Items 1 bis 6 des Aufgabentextes 1 in die jeweiligen Literalitätsstufen der IALS-Studie.

Tabelle 16: Aufgabenanalyse zu den Items 1 - 6 des Aufgabentextes 1 (nach Jordan 2011a, S. 199)

Aufgabe	IALS-Stufe	Mathematische Anforderungen	Rechenoperationen
1	Stufe 3	Flächenberechnung/Arithmetik: Rechteck/Multiplikation	$A = 15 \text{ m} * 20 \text{ m}$
2	Stufe 4	• Schritt: Umfangberechnung/ Arithmetik: Rechteck/ Multiplikation	$U = 2 * 15 \text{ m} + 2 * 20 \text{ m}$
		• Schritt: Arithmetik: Subtraktion	$70 \text{ m} - 2 \text{ m}$
3	Stufe 5	1. Schritt: Umfangberechnung/ Arithmetik: Rechteck/ Multiplikation	$U = 2 * 15 \text{ m} + 2 * 20 \text{ m}$
		2. Schritt: Berechnung der Mantelfläche	$M = (2 * 15 \text{ m} + 2 * 20 \text{ m}) * 3 \text{ m}$
		3. Schritt: Prozentrechnung: Prozentwert berechnen	$P = \frac{210 \text{ m}^2 * 20}{100}$
		4. Schritt: Arithmetik: Subtraktion	$210 \text{ m}^2 - 42 \text{ m}^2$
4	Stufe 2	Arithmetik: Multiplikation	$15 * 600$
5	Stufe 2/3	Arithmetik: Multiplikation	$58 * 6,5$
6	Stufe 2/3	Arithmetik: Multiplikation	$176 * 1,3$

Beim zweiten Aufgabentext handelt es sich um eine Kombination aus einem kontinuierlichen Text, einer Tabelle und einem Diagramm. In diesem Text wird eine Arbeitswoche eines Handelsvertreters beschrieben. Die Tabelle und das Diagramm dienen dazu, die gefahrenen Kilometer und den entsprechenden Kraftstoffverbrauch berechnen zu können. Der zweite Aufgabentext umfasst insgesamt drei Aufgaben (vgl. Jordan 2011a, S. 201). Bei Aufgabe 7 muss bestimmt werden, wie viele Kilometer Herr Hertelt am Montag gefahren ist. Der Proband muss aus dem Text die Informationen herausarbeiten, dass Herr Hertelt am Montagmorgen von seiner Wohnung aus Köln nach Bremen und abends von Bremen weiter nach Berlin fährt. Die

jeweiligen Entfernungen müssen anschließend aus der Tabelle in Form von Kilometerangaben entnommen und addiert werden (vgl. Jordan 2011a, S. 201). Aufgabe 8 unterscheidet sich von den Anforderungen nur geringfügig von Aufgabe 7. Aufgabe 8 ist schwieriger, weil die benötigten Zahlen dem Text nicht direkt zu entnehmen sind. Hat der Proband die Entfernungen schlussfolgernd aus der Tabelle bestimmen können, muss er diese addieren, um die Aufgabe zu lösen (vgl. Jordan 2011a, S. 202). Bei Aufgabe 9 muss der Proband den Fließtext, die Tabelle und das Diagramm hinzuziehen, um den Kraftstoffverbrauch für die Wegstrecke zwischen München und Stuttgart zu ermitteln. Der Proband muss zuvor aus dem Text die Art des Motors entnehmen, um anschließend die relevanten Informationen zum Kraftstoffverbrauch aus dem Diagramm ablesen zu können (vgl. Jordan 2011a, S. 202). Tabelle 17 verdeutlicht die Eingruppierung der Items 7 bis 9 des Aufgabentextes 2 in die jeweiligen Literalitätsstufen der IALS-Studie.

Tabelle 17: Aufgabenanalyse zu den Items 7 - 9 des Aufgabentextes 2 (nach Jordan 2011a, S. 201)

Aufgabe	IALS-Stufe	Mathematische Anforderungen	Rechenoperationen
7	Stufe 3	Arithmetik: Addition	$315 + 375$
8	Stufe 4	Arithmetik: Addition	$184 + 488$
9	Stufe 5	• Schritt: Tabelle lesen	210 km
		• Schritt: Diagramm lesen	$8 \frac{L}{100 km}$
		• Schritt: Arithmetik: Multiplikation und Division	$8L \frac{210 km}{100 km}$

Der dritte Aufgabentext umfasst wiederum einen Fließtext, eine Tabelle und ein Diagramm. Im Fließtext wird die Anforderung formuliert, herauszufinden, wie viel Alkohol in einem Cocktail enthalten ist. Der Aufgabentext umfasst insgesamt vier Items (vgl. Jordan 2011a, S. 202). Bei Aufgabe 10 muss bestimmt werden, wie viel ml Cocktail Frau Wernoth getrunken hat. Die Mengenangaben in ml für die einzelnen Cocktailgrößen sind dem Text direkt zu entnehmen und müssen anschließend addiert werden. Daher ist diese Aufgabe als einfach einzustufen (vgl. Jordan 2011a, S. 203).

Bei Aufgabe 11 muss der Proband den Fruchtsaftanteil eines „Mai Tai“ bestimmen. Hierfür muss der Proband die Tabelle interpretieren können, die die Zutaten für die Herstellung von Cocktails angibt. Hat der Proband die Fruchtsaftanteile aus der Tabelle herausgearbeitet, müssen diese miteinander addiert werden (vgl. Jordan 2011a, S. 203). Bei Aufgabe 12 wird gefordert, dass der Proband den Alkoholgehalt von Rum in Prozent berechnet. Zunächst muss der Proband den entsprechenden Alkoholgehalt von Rum dem Diagramm entnehmen und dann in einen Prozentwert umwandeln (vgl. Jordan 2011a, S. 203 f.). Bei Aufgabe 13 muss der Proband bestimmen, wie viel ml Alkohol ein kleiner „Tequila Sunrise“ beinhaltet. Hierfür muss der Proband Informationen aus dem Fließtext, der Tabelle und dem Diagramm miteinander verknüpfen, wobei mehrere Rechenoperationen benötigt werden. Diese Aufgabe stellt im TeMaTex das schwerste Item dar (vgl. Jordan 2011a, S. 203 f.). Tabelle 18 verdeutlicht die Eingruppierung der Items 10 bis 13 des Aufgabentextes 3 in die jeweiligen Literalitätsstufen der IALS-Studie.

Tabelle 18: Aufgabenanalyse zu den Items 10 - 13 des Aufgabentextes 3 (nach Jordan 2011a, S. 203)

Aufgabe	IALS-Stufe	Mathematische Anforderungen	Rechenoperationen
10	Stufe 1	Arithmetik: Addition bzw. Verdopplung	$300 + 300$ oder $2 \cdot 300$
11	Stufe 3/4	Bruchrechnung: Addition dreier Stammbrüche	$\frac{1}{6} + \frac{1}{3} + \frac{1}{20}$
12	Stufe 3	Prozentrechnung/Bruchrechnung: Umrechnung einer Bruch- in eine Prozentzahl	$\frac{2}{5} = \frac{40}{100} \Rightarrow 40\%$
		Alternativ: Arithmetik: Division und Interpretation des Ergebnisses als Prozentzahl	$2 : 5 = 0,4 \Rightarrow 40\%$
13	Stufe 5	1. Schritt: Bruchrechnung: Multiplikation zweier Brüche	$\frac{1}{4} \cdot \frac{1}{5}$ und $\frac{1}{6} \cdot \frac{2}{5}$
		2. Schritt: Bruchrechnung: Addition zweier Stammbrüche	$\frac{1}{20} + \frac{1}{15}$
		3. Schritt: Bruchrechnung: Multiplikation einer natürlichen Zahl mit einem Bruch	$300 \cdot \frac{7}{60}$

Die nachfolgenden Aufgaben sind Textaufgaben und „können nicht von der Struktur in das IALS-Konzept eingegliedert werden, weil bei diesen Aufgaben keine irrelevanten Informationen (bis auf zwei Aufgaben) im Aufgabentext [hinzugefügt] wurden. Daher kann an dieser Stelle nur eine Einteilung aufgrund der Schwierigkeit der Aufgaben vorgenommen werden“ (Jordan 2011a, S. 204). Bei Aufgabe 14 soll berechnet werden, nach wie vielen Arbeitstagen der Lagerbestand am zweiten Standort aufgebraucht ist. Insgesamt müssen die Probanden zwei verschiedene Rechenoperationen durchführen (vgl. Jordan 2011a, S. 205). Tabelle 19 verdeutlicht die Eingruppierung des Items 14 des Aufgabentextes 4 in die entsprechende Literalitätsstufe der IALS-Studie.

Tabelle 19: Aufgabenanalyse zum Item 14 des Aufgabentextes 4 (nach Jordan 2011a, S. 205)

Aufgabe	IALS-Stufe	Mathematische Anforderungen	Rechenoperationen
14	Stufe 4	I. Schritt: Arithmetik: Multiplikation	$12 * 200$
		II. Schritt: Arithmetik: Division	$2400 : 400$

Bei Aufgabe 15 muss der Proband den funktionalen Zusammenhang zwischen der zurückgelegten Entfernung und dem Preis erkennen. Diese Aufgabe ist ohne eine Rechenoperation lösbar (vgl. Jordan 2011a, S. 205). Zum Bewältigen von Aufgabe 16 ist das Lösen einer linearen Gleichung notwendig (vgl. Jordan 2011a, S. 205 f.). Tabelle 20 verdeutlicht die Eingruppierung der Items 15 und 16 des Aufgabentextes 5 in die jeweiligen Literalitätsstufen der IALS-Studie.

Tabelle 20: Aufgabenanalyse zu den Items 15+16 des Aufgabentextes 5 (nach Jordan 2011a, S. 205)

Aufgabe	IALS-Stufe	Mathematische Anforderungen	Rechenoperationen
15	Stufe 3	Interpretation eines funktionalen Zusammenhangs	
16	Stufe 5	Algebra: Lösen linearer Gleichungen	$3 + 1,5x = 15,75$

Bei Aufgabe 17 muss der Proband die Länge der Leiter anhand des Satzes des Pythagoras berechnen. In der Aufgabenstellung werden keine Hinweise zur erforderlichen Rechenoperation gegeben, was das Lösen der Aufgabe erschwert (vgl. Jordan 2011a, S. 206). Tabelle 21 verdeutlicht die

Eingruppierung des Items 17 des Aufgabentextes 6 in die entsprechende Literalitätsstufe der IALS-Studie.

Tabelle 21: Aufgabenanalyse zum Item 17 des Aufgabentextes 6 (nach Jordan 2011a, S. 206)

Aufgabe	IALS-Stufe	Mathematische Anforderungen	Rechenoperationen
17	Stufe 5	1. Schritt: Satz des Pythagoras: Aufstellen der Formel	$l^2 = 3^2 + 4^2$
		2. Schritt: Auflösen der Gleichung	$l = \sqrt{3^2 + 4^2}$
		3. Schritt: Radizieren von Quadratzahlen	$\sqrt{25}$

Bei Aufgabe 18 wird danach gefragt, wie viele Lötkolben in der 34. KW hergestellt werden. Der Aufgabentext enthält viele irrelevante Informationen, denen der Proband die richtigen entnehmen muss. Anschließend kann er durch eine einfache Rechenoperation die Aufgabe lösen (vgl. Jordan 2011a, S. 207). Tabelle 22 verdeutlicht die Eingruppierung von Item 18 des Aufgabentextes 7 in die entsprechende Literalitätsstufe der IALS-Studie.

Tabelle 22: Aufgabenanalyse zum Item 18 des Aufgabentextes 7 (nach Jordan 2011a, S. 207)

Aufgabe	IALS-Stufe	Mathematische Anforderungen	Rechenoperationen
18	Stufe 3	Arithmetik: Subtraktion	$634 - 409$

7.2 Gütekriterien des Tests

Im Folgenden werden die Gütekriterien Objektivität, Reliabilität und Validität in Bezug auf den TeMaTex erläutert.

7.2.1 Objektivität

Durchführungsobjektivität: Im Blick auf den vorliegenden Test wird die Durchführungsobjektivität auf verschiedene Weisen erreicht: Im Testmanual wird ein Instruktionsblatt zur Durchführung des Tests vorgegeben, welches den Probanden vor der Durchführung des Tests vom Testleiter vorgelesen wird. Somit werden die Probanden über den Ablauf informiert. Die Testbögen sind standardisiert, sodass für jede Testperson die gleichen Testbedingungen bestehen. Die Testzeit beträgt inklusive Testinstruktion durch den Testleiter 45 Minuten. Es sind keine Hilfsmittel wie Taschenrech-

ner erlaubt. Das Testmaterial besteht aus einem Testheft und einem Lösungsblatt. Die Probanden tragen die Lösungen zu den Testaufgaben aus dem Testheft ausschließlich in das vorgesehene Lösungsblatt ein. Hierfür liegen farbig hinterlegte Felder auf dem Lösungsblatt vor, in welche die Zahlen oder der Beträge eingetragen werden müssen (vgl. Jordan 2011a, S. 234).

Auswertungsobjektivität: Die Auswertungsobjektivität wird im vorliegenden Test durch folgende Regeln erreicht: Löst ein Proband ein Item richtig, wird dies mit einem Punkt bewertet. Es gelten nur diejenigen Antworten als richtig, welche in das vorgesehene Lösungsfeld in das Lösungsblatt eingetragen werden. Die in die Lösungsfelder eingetragenen Lösungen werden im Anschluss in eine Excel-gestützte Eingabemaske eingetragen, wo die Rohdaten automatisch umcodiert werden – zu richtig oder falsch beantworteten Aufgaben (vgl. Jordan 2011a, S. 34 f.).

Interpretationsobjektivität: Die Interpretationsobjektivität wird in diesen Zusammenhang wie folgt erreicht: Es liegen Prozentrangnormen für Schüler mit Haupt-, Real- und Gymnasialabschluss vor, welche zum Vergleich mit den erhobenen Daten herangezogen werden können (vgl. Jordan 2011a, S. 235).

7.2.2 Reliabilität

Retest-Reliabilität: In die Retest-Reliabilitätsanalyse des vorliegenden Tests wurden die Ergebnisse von insgesamt $n = 94$ einbezogen. Die Testwiederholung erfolgte 2 bis 4 Monate nach der ersten Erhebung. Der Reliabilitätskoeffizient in Bezug auf die Retest-Reliabilität des Tests beträgt $r = 0,780$ (vgl. Jordan 2011a, S. 235 f.).

Paralleltest-Reliabilität: Zum vorliegenden Test liegen keine Ergebnisse zur Paralleltest-Reliabilität vor.

Testhalbierungs-Reliabilität: Zur Aufteilung in zwei Testhälften wurde die Even-Odd-Methode gewählt. Insgesamt wurden $n = 2567$ Probanden in die

Analyse einbezogen. Die Testhalbierungs-Reliabilität des vorliegenden Tests liegt bei $r = 0,876$ (vgl. Jordan 2011a, S. 236 f.).

Innere Konsistenz: Aufgrund der dichotomen Datenstruktur wurde die Kuder-Richardson-Formel 20 zur Berechnung des Koeffizienten verwendet. Die innere Konsistenz des vorliegenden Tests liegt bei $r = 0,852$ (vgl. Jordan 2011a, S. 236).

7.2.3 Validität

Inhaltsvalidität: Im Rahmen der Itemanalyse wurden 18 Items und deren Lösungsschritte detailliert beschrieben, um aufzuzeigen, dass alle Items das Konstrukt erfassen. Auf der Gesamttestebene wurde sichergestellt, dass die mathematischen Grundlagenbereiche wie Arithmetik, Prozentrechnung und Bruchrechnung durch jeweils mindestens ein Item abgefragt wurden (vgl. Jordan 2011a, S. 237).

Konstruktvalidität: Für den TeMaTex wurde mithilfe der Software WINMIRA die Modellgeltung unter Verwendung des Bootstrap-Verfahrens nachgewiesen. Zur ausführlichen Beschreibung des Verfahrens zur Ermittlung der Konstruktvalidität des TeMaTex sei auf die Arbeit von Jordan 2011 (vgl. Jordan 2011a, S. 243 ff.) verwiesen.

Kriteriumsvalidität: Die erreichten Korrelationskoeffizienten können Kapitel 7.3 entnommen werden. Zur ausführlichen Beschreibung des Verfahrens zur Ermittlung der Kriteriumsvalidität des TeMaTex sei auf die Arbeit von Jordan 2011 (Jordan 2011a, S. 237 ff.) verwiesen.

7.3 Zusammenfassende Betrachtung des siebten Kapitels

Im Rahmen des Testverfahrens erhält jeder Proband ein Testheft und einen Bearbeitungsbogen. Das Testheft wird nicht beschrieben, da die Endergebnisse auf dem Bearbeitungsbogen eingetragen werden müssen. Die Auswertung der Daten erfolgt mit einer Excel-Datei. Hierfür trägt der Testleiter die Namen und Testantworten der Probanden in die Excel-Datei ein. Die Auswertung erfolgt dann automatisch (vgl. Jordan & Stein 2011b).

Die Zuverlässigkeit des TeMaTex wurde anhand der Testhalbierungs-Reliabilität, der Re-Test-Reliabilität und mittels Cronbachs Alpha bestimmt. Die Testhalbierungs-Reliabilität liegt bei $r = 0,908$ ($n = 124$) respektive $r = 0,876$ ($n = 2.567$), die Re-Test-Reliabilität bei $r = 0,780$ und die Messung der inneren Konsistenz nach Cronbachs Alpha ergab ein r von $0,880$ ($n = 124$) respektive $r = 0,852$ ($n = 2.567$). Im Rahmen der Überprüfung der Gültigkeit des Tests wurde bei der kriterienbezogenen Validität eine Korrelation mit dem Wortschatztest WT von $r = 0,50$ ($n = 252$), mit dem Zahlenfolgetest ZT eine Korrelation von $r = 0,61$ ($n = 242$), mit dem Intelligenztest CFT-R 20 eine Korrelation von $r = 0,60$ ($n = 241$) und mit dem Lesetest LTB-3 eine Korrelation von $r = 0,56$ ($n = 1.969$) ermittelt. Die Überprüfung der Konstruktvalidität erfolgte über eine erfolgreiche Faktoren- und Rasch-Analyse zum Nachweis der Konstruktvalidität des Tests. Die Durchführungsobjektivität des TeMaTex wird laut Jordan und Stein (2011b) gewährleistet, indem der Testleiter jeweils ein Instruktionsblatt vorliest, mit dem die Probanden über den Test informiert werden. Die Testbögen des TeMaTex sind standardisiert und für alle Probanden gelten dieselben Durchführungsbedingungen (vgl. Jordan & Stein 2011b).

Die Auswertungsobjektivität des Tests wird erreicht, indem jede korrekte Antwort zu den Items 1 bis 18 einen Punkt erhält. Als korrekt gelten diejenigen Antworten, welche in das jeweilige Antwortfeld auf dem Bearbeitungsbogen eingetragen werden. Die Nebenrechnungen bei den Aufgaben werden nicht bewertet. Die Excel-Auswertungssoftware ermittelt automatisch, ob ein Item richtig oder falsch beantwortet wurde. Die Interpretationsobjektivität wird dadurch erreicht, dass im Rahmen der Testentwicklung verschiedene Schulformen getestet wurden und für diese unter anderem Prozentrangnormen angegeben werden können. So kann die Leistung eines Probanden anhand seiner Leistungsgruppe bewertet werden (vgl. Jordan & Stein 2011b). Die Normierungsstichprobe bestand aus 2567 Probanden. Für folgende Gruppen können Prozentränge angegeben werden: Schüler (9. und 10 Klasse) der Haupt- und Realschule sowie Schüler des Gymnasi-

ums (9. und 10. Klasse). Darüber hinaus liegen Prozentrangnormen für Berufsschüler vor. Der TeMaTex ist seit 2011 in der Anwendung (vgl. Jordan & Stein 2011b).

III. Empirischer Teil 1

Im empirischen Teil 1 werden die Ergebnisse der Untersuchung in Bezug auf den TeMaTex deskriptiv und inferenzstatistisch dargestellt.

8 Statistische Analysen in Bezug auf den TeMaTex

Im folgenden Kapitel 8.1 werden die Ergebnisse der deskriptiven Analyse vorgestellt. Hierbei wird auf die Zusammensetzung der Stichprobe eingegangen sowie auf die Antworthäufigkeiten der Probanden, unterteilt nach Schulabschlüssen. Die Darstellung der Ergebnisse der inferenzstatistischen Analyse findet in Kapitel 8.2 statt. In diesem Kapitel wird die Modellentwicklung mithilfe der verschiedenen Statistikprogramme erläutert. Abschließend wird in Kapitel 8.3 eine Zusammenfassung des gesamten Kapitels gegeben.

8.1 Deskriptive Analyse

Das Alter der Probanden in der Stichprobe ($n = 411$) liegt zwischen 16 und 56 Jahren. Der Median liegt bei 22 Jahren. Von den 411 Probanden haben 38 keine Altersangaben gemacht. Im Rahmen der Datenerhebung wurden zudem Angaben zum Schulabschluss der Probanden erhoben. Tabelle 23 gibt einen Überblick über die Verteilung der Schulabschlüsse in der Stichprobe.

Tabelle 23: Verteilung Schulabschlüsse - empirischer Teil I

Hauptschulabschluss	Realschulabschluss	Fachhochschulreife/Abitur	Keine Angaben
125	182	67	37

Der Realschulabschluss stellt mit 44,2 % den am häufigsten angegebenen Schulabschluss in der Stichprobe dar, gefolgt vom Hauptschulabschluss mit 30,3 %. Probanden mit Fachhochschulreife beziehungsweise mit Abitur machen hingegen nur 16,3 % aus. Circa 9,0 % machten keine Angaben zu ihrem Schulabschluss.

Im Folgenden werden die in der Untersuchung verwendeten 18 Items des TeMaTex in Bezug auf ihre Antworthäufigkeiten beschrieben. In der zweiten Spalte von Tabelle 24 werden die Items und in der letzten Spalte die Anzahl der getesteten Probanden angegeben. Die Spalten 4 und 5 geben die Häufigkeiten der gegebenen Antworten zu den jeweiligen Items an (0 = keine oder falsche Antwort, 1 = richtige Antwort). Aus der Tabelle geht hervor, wie häufig eine Aufgabe von den Testpersonen gelöst oder nicht gelöst beziehungsweise nicht bearbeitet wurde. Item 17 (Aufgabe 17 des Aufgabentextes 6: Der Bungalow) wurde von den Probanden am häufigsten nicht gelöst oder nicht beantwortet (0 = 392). Die Probanden beantworteten dieses Item lediglich 19-mal richtig. Item 10 (Aufgabe 10 des Aufgabentextes 3: Die Cocktailbar) wurde am häufigsten richtig gelöst (1 = 270).

Tabelle 24: Antworthäufigkeiten der 18 Items des TeMaTex - empirischer Teil I

Item labels and sample frequencies:					
no.	n of label	categories	0	1	N
1	VAR1	2	178	233	411
2	VAR2	2	242	169	411
3	VAR3	2	351	60	411
4	VAR4	2	198	213	411
5	VAR5	2	260	151	411
6	VAR6	2	294	117	411
7	VAR7	2	155	256	411
8	VAR8	2	187	224	411
9	VAR9	2	328	83	411
10	VAR10	2	141	270	411
11	VAR11	2	381	30	411
12	VAR12	2	322	89	411
13	VAR13	2	390	21	411
14	VAR14	2	318	93	411
15	VAR15	2	274	137	411
16	VAR16	2	336	75	411
17	VAR17	2	392	19	411
18	VAR18	2	335	76	411

Im weiteren Verlauf wurden die Häufigkeiten der empirisch erreichten Gesamtpunktzahlen nach den unterschiedlichen Schulabschlüssen aufgeteilt.

Tabelle 25 können diese entnommen werden. In der linken Spalte der Tabelle stehen die theoretisch erreichbaren Gesamtpunktzahlen (0 bis 18 Punkte). In der zweiten, dritten und vierten Spalte werden die Häufigkeiten der jeweils erreichten Gesamtpunktzahlen genannt.

Tabelle 25: Verteilung der Gesamtpunktwerte nach Schulabschluss - empirischer Teil I

Ge- samt- punkt- werte	Hauptschulab- schluss	Realschulab- schluss	Fachhochschulab- schluss/Abitur
0	20	11	1
1	17	9	1
2	14	6	3
3	14	24	3
4	16	24	3
5	14	14	4
6	8	19	8
7	10	20	2
8	5	20	7
9	2	11	11
10	0	6	2
11	1	10	4
12	0	1	4
13	1	3	4
14	2	1	3
15	0	0	2
16	1	3	4
17	0	0	1
18	0	0	0
	n = 125	n = 182	n = 67

Abbildung 9 zeigt einen Vergleich zwischen den Personenlokalisationen und den Itemlokalisationen. Die durchschnittlichen Personenfähigkeiten liegen mit einem Wert von -1,279 deutlich unterhalb der durchschnittlichen Itemschwierigkeiten von 0,000. Bei ca. 90 Personen ist die Fähigkeit niedriger als die geringste Itemschwierigkeit. Es zeigt sich auch, dass der Test keine Items umfasst, die nicht gelöst wurden.

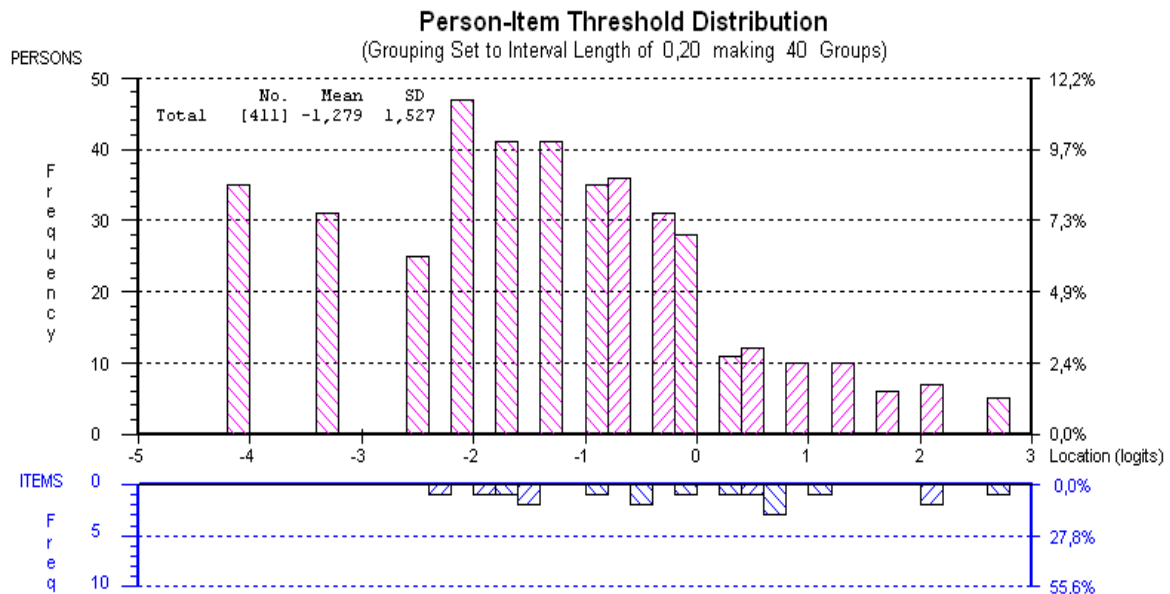


Abbildung 9: RUMM 2030: Verteilung Personenfähigkeit und Itemschwierigkeit - empirischer Teil I

8.2 Inferenzstatistische Analyse

Im Rahmen der schließenden Statistik werden die in Kapitel 1.2 formulierten Hypothesen überprüft. Grundlage für die Berechnungen sind Daten auf Intervallskalenniveau. Mithilfe der Parameterschätzung im dichotomen Rasch-Modell erlangen die Daten bei Modellpassung das entsprechende Skalenniveau. Beim Vorliegen der spezifischen Objektivität sind die Ergebnisse stichprobenunabhängig (vgl. Moosbrugger 2007d, S. 230 f.).

8.2.1 Ergebnisse der Modellentwicklung mit dem Rasch-Modell

Um die Konstruktvalidität des Diagnoseinstrumentes TeMaTex ermitteln zu können, wurde die Rasch-Analyse durchgeführt. In die Rasch-Analyse wurden insgesamt $n = 411$ Probanden einbezogen (Tabelle 26).

Tabelle 26: Allgemeine Daten zur Rasch-Analyse - empirischer Teil I

number of persons:	411
number of items:	18
number of classes:	1
max. number of iterations:	250
accuracy criterion:	0.0005
random start value:	4321

Zur Analyse der Daten wurden die Statistikprogramme RUMM 2030, WINMIRA und eRm eingesetzt. Die Auswertungen der verschiedenen Softwareprogramme befinden sich auf der Daten-CD unter den entsprechenden Ordnern. Zu Beginn werden die Daten mit RUMM 2030 ausgewertet und anschließend mit den Ergebnissen der Analysen aus WINMIRA und eRm abgeglichen. Aufgrund der dichotomen Datenstruktur (0 = falsche Lösung oder die Aufgabe wurde nicht bearbeitet und 1 = korrekte Lösung) wurde zur Analyse das 1-PL-Rasch-Modell gewählt. Der TeMaTex umfasst 18 Items, woraus sich 262.144 mögliche theoretische Antwortmuster erstellen lassen (Tabelle 27). Empirisch liegen allerdings nur 300 Antwortmuster vor, sodass die Voraussetzungen für eine χ^2 -Asymptotik des globalen Modellgeltungstests nicht vorliegen. Hierfür müsste jedes Antwortmuster mindestens einmal auftreten (vgl. Rost 2004).

Tabelle 27: Überblick Rasch-Analyse - empirischer Teil I

saturated likelihood:	-2212.1230
number of different patterns:	300
number of possible patterns:	262144

8.2.1.1 Rasch-Analyse mit RUMM 2030

Zunächst sollen Angaben zur Teststruktur, zur Stichprobe und zu den Parameterschätzungen gemacht werden (Tabelle 28).

Tabelle 28: RUMM 2030: Überblick Rasch-Analyse mit RUMM 2030 - empirischer Teil I

TEST STRUCTURE
No. of Items: 18
No. of Categories Equal across Items: set at 2 categories per item
Score Range [All Items]: 17
CALIBRATING SAMPLE
No. of Persons:
* entered Project: 411
* invalid records: 0
* extreme scores: 35
* valid scores: 376 [available for analysis]
Missing data detected: None
ESTIMATION DETAILS
Item Converge Limit: 0,00010
Iterations [Items]: 4 [all 18 parameters converged]
Iterations [Persons]: -1
Lowest score: 1
Highest score: 17
Person Converge Limit: 0,01000
Person Estimation: Weighted Maximum Likelihood method
Extreme Pers. Criterion: 0,220

RUMM 2030 bietet durch sein „Pairwise“-Algorithmus unterstütztes Schätz-Verfahren, eine Möglichkeit an, extreme Antwortmuster auszuschließen (vgl. Brühl 2012, S. 43). „Der pairwise-Algorithmus ist im Vergleich zum Weighted Likelihood Estimate (WLE)-Algorithmus das strengere Verfahren, da er alle diejenigen Fälle aus der Berechnung ausschließt, die Extremwerte aufweisen“ (Brühl 2012, S. 43). Zum Zeitpunkt der Analyse liegen keine ungültigen oder fehlenden Werte vor. Insgesamt wurden daher 376 gültige Fälle berücksichtigt (Tabelle 28). Für die Schätzungen der Itemparameter wurden 4 Iterationen (Wiederholungen) benötigt. Im Folgenden werden die ICC der Items 10 (Aufgabe 10 des Aufgabentextes 3: Die Cocktailbar), 13 (Aufgabe 13 des Aufgabentextes 3: Die Cocktailbar) und 17 (Aufgabe 17 des Aufgabentextes 6: Der Bungalow) dargestellt. Bei der Betrachtung der einzelnen ICC der 18 Items fällt auf, dass, wie Tabelle 24 zu entnehmen ist, Item 10 mit einer Item-Schwierigkeit von -2,265 das leichteste zu lösende Item darstellt und Item 17 mit einer Item-Schwierigkeit von 2,667 das schwierigste.

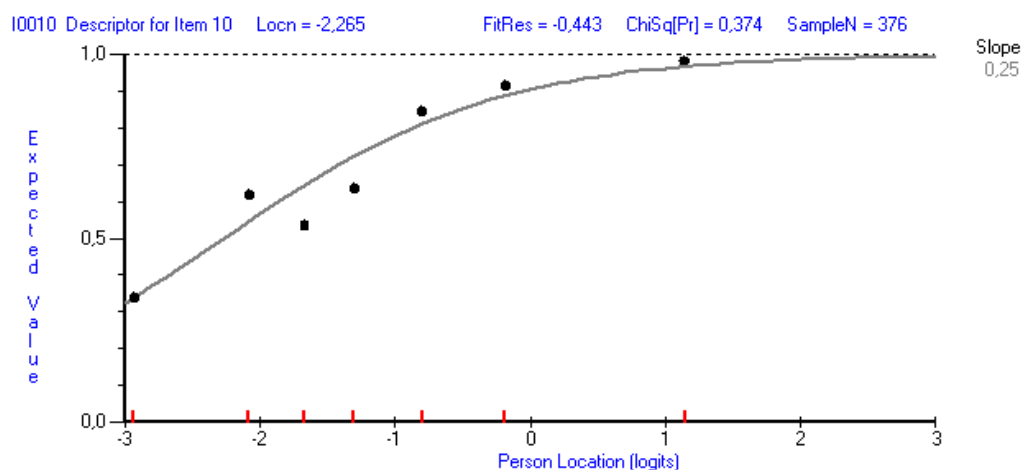


Abbildung 10: RUMM 2030: Item Characteristic Curve von Item 10 - empirischer Teil I

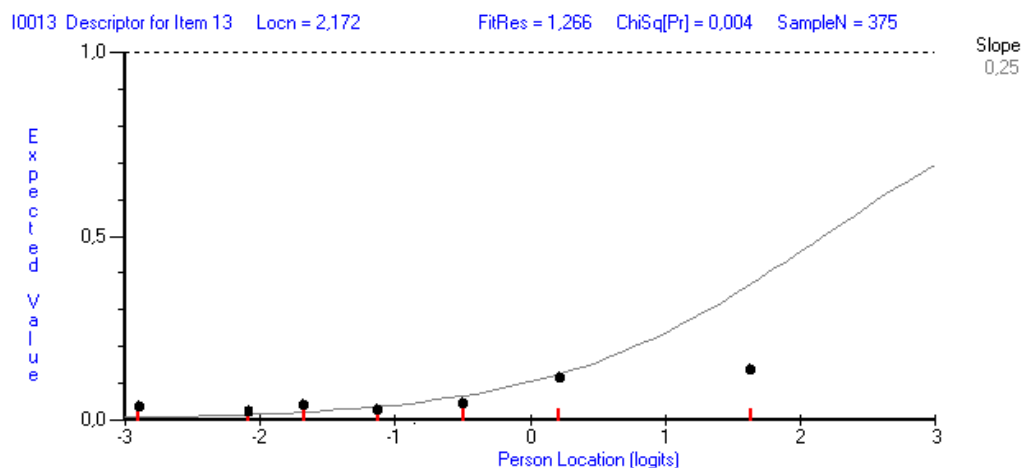


Abbildung 11: RUMM 2030: Item Characteristic Curve von Item 13 - empirischer Teil I

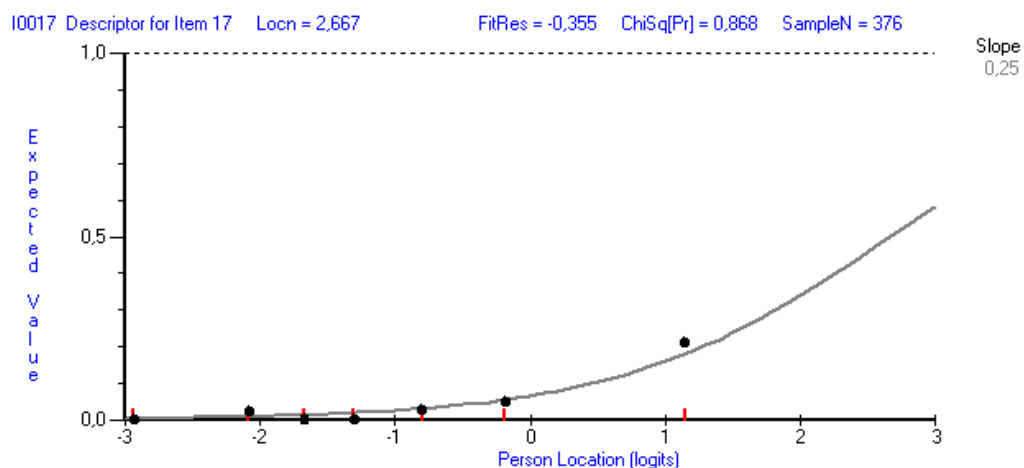


Abbildung 12: RUMM 2030: Item Characteristic Curve von Item 17 - empirischer Teil I

Im Rahmen der Modellentwicklung werden die Fit-Residuen der 18 Items genauer betrachtet. Hierbei werden die beobachteten mit den erwarteten Werten verglichen. Die Differenzen werden in Fit-Residuen ausgedrückt, wobei Fit-Residuen-Werte von mehr als ± 2 als signifikant zu betrachten sind. Diese Items müssen aus dem Test eliminiert werden, da eine Modellverletzung vorliegt (Tabelle 29: p -Wert = 0,000865).

Tabelle 29: RUMM 2030: Übersicht Item-Person und Item-Trait Interaction - empirischer Teil I

ITEM-PERSON INTERACTION				
=====				
	ITEMS		PERSONS	
	Location	Fit Residual	Location	Fit Residual

Mean:	0,000	-0,223	-1,279	-0,251
SD:	1,459	1,158	1,542	0,740
Skewness:		0,022		0,772

Kurtosis:	-1,271	0,687
Correlation:	0,000	0,085
Complete data DF =	0,942	

=====		
ITEM-TRAIT INTERACTION		RELIABILITY INDICES

Total Item Chi Squ:	160,035	Separation Index: 0,77314
Total Deg of Freedom:	108,000	Cronbach Alpha: 0,83736
Total Chi Squ Prob:	0,000865	

=====		
LIKELIHOOD-RATIO TEST		POWER OF TEST-OF-FIT

Chi Squ	Power is GOOD	
Degrees of Freedom	[Based on SepIndex of 0,77314]	
Probability		

Tabelle 30 lässt sich entnehmen, dass Item 9 (Aufgabe 9 des Aufgabentextes 2: Der Handelsvertreter) mit -2,161 über dem kritischen Wert liegt.

Tabelle 30: RUMM 2030: Übersicht Individual Item-Fit Werte - empirischer Teil I

INDIVIDUAL ITEM-FIT - Serial Order								

Seq	Item	Location	SE	Residual	DF	ChiSq	DF	Prob

1	I0001	-1,658	0,121	1,695	353,22	8,577	6	0,198773
2	I0002	-0,799	0,120	1,081	353,22	12,902	6	0,044618
3	I0003	1,121	0,160	-1,665	353,22	8,609	6	0,196765
4	I0004	-1,466	0,120	-1,095	353,22	5,614	6	0,467781
5	I0005	-0,547	0,122	0,354	353,22	15,128	6	0,019287
6	I0006	-0,058	0,128	0,375	353,22	4,557	6	0,601789
7	I0007	-1,939	0,124	1,378	353,22	4,452	6	0,615752
8	I0008	-1,529	0,120	-0,046	353,22	11,456	6	0,075256
9	I0009	0,609	0,143	-2,161	353,22	9,761	6	0,135085
10	I0010	-2,280	0,129	-0,482	353,22	7,039	6	0,317278
11	I0011	2,040	0,205	-0,924	353,22	1,733	6	0,942542
12	I0012	0,414	0,138	-1,650	353,22	9,677	6	0,138911
13	I0013	2,172	0,213	1,266	353,22	19,229	6	0,003795
14	I0014	0,349	0,136	-1,633	353,22	6,780	6	0,341638
15	I0015	-0,435	0,123	-0,360	353,22	15,145	6	0,019159
16	I0016	0,681	0,145	-0,393	353,22	9,488	6	0,147943
17	I0017	2,683	0,251	-0,348	353,22	4,008	6	0,675642
18	I0018	0,642	0,144	0,597	353,22	5,879	6	0,436867

Dieses Item wurde aus dem Datensatz eliminiert. Danach fand eine erneute Überprüfung der Modellgeltung mit dem reduzierten Itemsatz statt. Tabelle

31 zeigt die Ergebnisse der wiederholten Modellprüfung. Trotz der auf Fit-Residuen basierenden Itemreduktion liegt eine Modellverletzung vor.

Tabelle 31: RUMM 2030: Ergebnisse Fit-Residuen basierender Itemreduktion - empirischer Teil I

ITEM-TRAIT INTERACTION		RELIABILITY INDICES
Total Item Chi Squ:	162,073	Separation Index: 0,76180
Total Deg of Freedom:	102,000	Cronbach Alpha: 0,82549
Total Chi Squ Prob:	0,000146	

Im Rahmen der Modellprüfung gibt es bei RUMM 2030 eine weitere Möglichkeit, durch Itemreduktion eine mögliche Modellanpassung zu erreichen. Hierfür werden die χ^2 -Werte der jeweiligen Items zugrunde gelegt. Ist ein Wert statistisch signifikant ($p < 0,05$), wird das jeweilige Item aus dem Datensatz eliminiert und eine erneute Modellprüfung durchgeführt. Die Ergebnisse der wiederholten Modellprüfung zeigen eine Modellgeltung für die auf den χ^2 -Werten basierende Itemreduktion an (Tabelle 32).

Tabelle 32: RUMM 2030: Ergebnisse χ^2 basierende Itemreduktion – empirischer Teil I

ITEM-TRAIT INTERACTION		RELIABILITY INDICES
Total Item Chi Squ:	88,340	Separation Index: 0,73277
Total Deg of Freedom:	84,000	Cronbach Alpha: 0,80967
Total Chi Squ Prob:	0,351731	

Es wurden insgesamt vier Items eliminiert. Dazu zählen (Tabelle 33) Item 2 und 5 (Aufgaben 2 und 5 des Aufgabentextes 1: Die Renovierungsfirma), Item 13 (Aufgabe 13 des Aufgabentextes 3: Die Cocktailbar) und Item 15 (Aufgabe 15 des Aufgabentextes 5: Die Taxifahrt).

Tabelle 33: RUMM 2030: Übersicht individual Item-Fit Werte der vier eliminierten Items -empirischer Teil I

INDIVIDUAL ITEM-FIT - Serial Order								
Seq	Item	Location	SE	Residual	DF	ChiSq	DF	Prob
1	I0001	-1,658	0,121	1,695	353,22	8,577	6	0,198773
2	I0002	-0,799	0,120	1,081	353,22	12,902	6	0,044618
3	I0003	1,121	0,160	-1,665	353,22	8,609	6	0,196765
4	I0004	-1,466	0,120	-1,095	353,22	5,614	6	0,467781
5	I0005	-0,547	0,122	0,354	353,22	15,128	6	0,019287
6	I0006	-0,058	0,128	0,375	353,22	4,557	6	0,601789

7	I0007	-1,939	0,124	1,378	353,22	4,452	6	0,615752
8	I0008	-1,529	0,120	-0,046	353,22	11,456	6	0,075256
9	I0009	0,609	0,143	-2,161	353,22	9,761	6	0,135085
10	I0010	-2,280	0,129	-0,482	353,22	7,039	6	0,317278
11	I0011	2,040	0,205	-0,924	353,22	1,733	6	0,942542
12	I0012	0,414	0,138	-1,650	353,22	9,677	6	0,138911
13	I0013	2,172	0,213	1,266	353,22	19,229	6	0,003795
14	I0014	0,349	0,136	-1,633	353,22	6,780	6	0,341638
15	I0015	-0,435	0,123	-0,360	353,22	15,145	6	0,019159
16	I0016	0,681	0,145	-0,393	353,22	9,488	6	0,147943
17	I0017	2,683	0,251	-0,348	353,22	4,008	6	0,675642
18	I0018	0,642	0,144	0,597	353,22	5,879	6	0,436867

RUMM 2030 bietet auch die Möglichkeit an, die Items auf das Vorliegen eines *Differential Item Functioning* (DIF) zu überprüfen. Die Ergebnisse der Überprüfung zeigen, dass die Items 3, 4, 10, 12 und 14 ein DIF aufweisen. Tabelle 34 gibt eine Zusammenfassung über vorhandene Differential Item Functioning der 18 Items.

Tabelle 34: RUMM 2030: Zusammenfassung Differential Item Functioning der 18 Items des TeMaTex - empirischer Teil I

SUMMARY of ANOVA for each Item

Item	Class Interval				Schulab				Schulab-x-CInt				Total DIF			
	MS	F	DF	p	MS	F	DF	p	MS	F	DF	p	MS	F	DF	p
I0001	0,51	15,149	5	0,004844	0,06	1,659	2	0,280189	-0,01	-0,190	6	**N/Sig	0,07	0,272	8	0,949541
I0002	0,85	1,004	5	0,498219	0,38	0,448	2	0,662471	0,21	0,243	6	0,942595	2,00	0,294	8	0,939020
I0003	0,93	408,529	5	0,000002	0,55	240,375	2	0,000011	0,32	139,721	6	0,000021	3,01	164,884	8	0,000013
I0004	1,87	1423,319	5	0,000000	0,78	595,445	2	0,000001	0,62	473,240	6	0,000001	5,30	503,791	8	0,000001
I0005	0,92	0,937	5	0,527700	0,35	0,360	2	0,714477	0,22	0,223	6	0,952390	2,01	0,257	8	0,956236
I0006	0,71	341,556	5	0,000003	0,01	2,960	2	0,141877	0,78	375,569	6	0,000002	4,71	282,417	8	0,000003
I0007	1,46	2,654	5	0,153888	1,46	2,640	2	0,164983	0,37	0,673	6	0,680962	5,14	1,164	8	0,453323
I0008	1,21	1,118	5	0,452735	1,87	1,724	2	0,269468	0,65	0,604	6	0,722293	7,66	0,884	8	0,583311
I0009	0,45	0,982	5	0,507561	1,54	3,344	2	0,119703	1,11	2,418	6	0,175602	9,77	2,649	8	0,149056
I0010	1,36	493,659	5	0,000001	0,46	166,634	2	0,000027	0,31	111,231	6	0,000037	2,75	125,082	8	0,000025
I0011	2,04	0,800	5	0,593894	2,34	0,920	2	0,456818	1,31	0,512	6	0,780357	12,52	0,614	8	0,742911
I0012	1,07	761,981	5	0,000000	0,09	64,459	2	0,000269	-0,03	-19,565	6	**N/Sig	0,02	1,441	8	0,357185
I0013	0,51	0,512	5	0,759732	0,38	0,385	2	0,698816	0,15	0,150	6	0,980579	1,65	0,209	8	0,974590
I0014	0,45	654,620	5	0,000000	0,31	448,417	2	0,000002	0,30	436,591	6	0,000001	2,43	439,548	8	0,000001
I0015	1,59	3,164	5	0,115960	0,61	1,220	2	0,370173	0,46	0,906	6	0,554228	3,97	0,985	8	0,532439
I0016	0,65	1,343	5	0,377043	0,76	1,574	2	0,294913	0,34	0,699	6	0,665565	3,53	0,918	8	0,565933
I0017	0,12	280,631	5	0,000004	0,01	15,645	2	0,007046	0,00	-2,429	6	**N/Sig	0,01	2,089	8	0,216491
I0018	0,06	280,583	5	0,000004	0,00	15,642	2	0,007049	0,00	-2,429	6	**N/Sig	0,00	2,089	8	0,216551
Item	Class Interval				Schulab				Schulab-x-CInt				Total DIF			

Im Rahmen weiterer Untersuchungen, in denen es darum geht, das mathematische Textverständnis von Altenpflegeschülern mit unterschiedlichen Schulabschlüssen mittels dieser 18 Items des TeMaTex zu erfassen, ist zu überlegen, ob die Items 3, 4, 10, 12 und 14 aus dem Test zu entfernen sind. Diese Items weisen eine DIF auf und sind somit für eine Differenzierung des latenten Merkmals in Abhängigkeit zum Schulabschluss wenig geeignet.

8.2.1.2 Rasch-Analyse mit WINMIRA

In der Spalte „Expected freq.“ wird die theoretisch erwartete Punkteverteilung der 411 Probanden unter Voraussetzung des Rasch-Modells dargestellt (Tabelle 35). Theoretisch müssten 30 Probanden keine und ein Proband alle Testaufgaben richtig beantwortet haben. In den nächsten Spalten sind die Schätzungen der Personenparameter durch die Maximum-Likelihood-Schätzung (MLE) und die Weighted-Likelihood-Schätzung (WLE) tabellarisch aufgelistet. Die Personenparameter anhand der WLE liegen zwischen $\theta_0 = -4.498$ bei einer Gesamtpunktzahl von 0 Punkten und $\theta_{18} = 4.652$ bei einer Gesamtpunktzahl von 18 Punkten. Für die Probanden, die die Hälfte (50 %) der Testaufgaben richtig gelöst haben, liegt ein Personenparameter von $\theta_9 = -0.029$ vor, welcher den Vorgaben des Rasch-Modells entspricht.

Tabelle 35: WINMIRA: Übersicht Personenparameter - empirischer Teil I

Raw-score	Expected freq.	MLE-estimate	std. error MLE	WLE-estimate	std. error WLE
0	30.56	*****	*****	-4.498	1.546
1	34.37	-3.607	1.068	-3.261	0.942
2	37.30	-2.769	0.806	-2.604	0.770
3	39.03	-2.210	0.700	-2.113	0.686
4	39.41	-1.761	0.644	-1.699	0.638
5	38.38	-1.370	0.610	-1.328	0.607
6	36.06	-1.012	0.589	-0.984	0.587
7	32.68	-0.674	0.575	-0.657	0.574
8	28.57	-0.348	0.568	-0.340	0.567
9	24.09	-0.028	0.566	-0.029	0.566
10	19.60	0.294	0.569	0.283	0.569
11	15.38	0.623	0.579	0.601	0.578
12	11.64	0.967	0.596	0.932	0.594
13	8.50	1.337	0.622	1.287	0.618
14	5.99	1.748	0.662	1.678	0.655
15	4.07	2.224	0.723	2.124	0.709

16	2.67	2.820	0.831	2.658	0.797
17	1.69	3.702	1.089	3.365	0.973
18	1.03	*****	*****	4.652	1.584

Der Mittelwert der Personenparameterschätzung liegt bei -1.326 (Tabelle 36). Der Anova-Reliabilitätskoeffizient (0.811) und der Reliabilitätskoeffizient nach Andrich (0.767) geben Hinweise auf eine ausreichend vorhandene innere Konsistenz. Der Normbereich der beiden Reliabilitätskoeffizienten liegt zwischen 0,8 und 0,9 (vgl. Bühner 2006, S. 140).

Tabelle 36: WINMIRA: Übersicht Mittelwert Personenparameterschätzung - empirischer Teil I

WLE estimates:
Mean = -1.326
Var = 2.603
stdev = 1.613
marginal error variance = 0.605/stdev = 0.778
anova reliability = 0.811
Andrichs reliability = 0.767

Die geschätzten Itemparameter für die 18 Testaufgaben liegen im Bereich von $\theta_{10} = -2.26$ und $\theta_{17} = 2.62$. Fünf Items sind in den Bereich der leichteren Items unter -1 und vier Items in den Bereich der schwereren Items über 1 anzusiedeln. Insgesamt neun Items sind im mittleren Schwierigkeitsbereich von -1 bis 1 anzuordnen (Tabelle 37).

Tabelle 37: WINMIRA: Übersicht Itemschwierigkeiten der Items - empirischer Teil I

item label	item location	threshold parameters
VAR1	-1.71941	
VAR2	-0.83963	
VAR3	1.06235	
VAR4	-1.44275	
VAR5	-0.58476	
VAR6	-0.06710	
VAR7	-2.04995	
VAR8	-1.59410	
VAR9	0.54359	
VAR10	-2.26216	
VAR11	2.04394	
VAR12	0.42543	
VAR13	2.49619	
VAR14	0.34960	
VAR15	-0.37889	
VAR16	0.71070	
VAR17	2.61780	
VAR18	0.68914	

Zur Bestimmung der lokalen Modellgeltung der 18 Items werden hierfür die Q-Indizes und die ZQ-Werte der jeweiligen Items analysiert (Tabelle 38).

Tabelle 38: WINMIRA: Übersicht Q-Indizes und ZQ-Werte - empirischer Teil I

Item label	Q-index	Zq	p(X>Zq)	
VAR1	0.1401	0.4039	0.34313	-...Q.!....+
VAR2	0.1510	0.6981	0.24256	-.Q..!....+
VAR3	0.1072	-1.5256	0.93645	-....!...Q+
VAR4	0.1086	-0.6583	0.74484	-....!.Q..+
VAR5	0.1539	0.7371	0.23052	-.Q..!....+
VAR6	0.1521	0.6104	0.27079	-.Q..!....+
VAR7	0.1528	0.9000	0.18406	-Q...!....+
VAR8	0.1284	-0.0196	0.50781	-....Q....+
VAR9	0.1029	-1.4400	0.92507	-....!...Q+
VAR10	0.1209	-0.0403	0.51609	-....Q....+
VAR11	0.1712	0.0334	0.48668	-...Q!....+
VAR12	0.1045	-1.3957	0.91860	-....!...Q+
VAR13	0.3299	3.0666	0.00108-!	Q....!....+
VAR14	0.1134	-0.9640	0.83247	-....!.Q..+
VAR15	0.1274	-0.2045	0.58100	-....Q....+
VAR16	0.1359	-0.2079	0.58233	-....Q....+
VAR17	0.1594	-0.6515	0.74265	-....!.Q..+
VAR18	0.1470	0.2269	0.41026	-...Q!....+
-?:p<0.05, +?:p>0.95				
-!:p<0.01, +!:p>0.99				

Aus Tabelle 38 geht hervor, dass alle Werte bis auf Item 13 (Aufgabe 13 des Aufgabentextes 3: Die Cocktailbar) im gültigen Akzeptanzbereich liegen. Item 13 weist zudem einen Item-Underfit auf.

Als Verfahren zur Überprüfung der globalen Modellgeltung werden der Likelihood-Ratio-Test und der χ^2 -Test eingesetzt. Der Cressie-Read-Wert und der Pearson χ^2 -Wert liegen bei $p = 1.0$, was auf eine Modellgeltung hindeutet (Tabelle 39).

Tabelle 39: WINMIRA: Übersicht Ergebnisse Modellgeltungstest - empirischer Teil I

Goodness of fit statistics:

	estimated model	saturated model
Log-Likelihood:	-3346.28	-2212.12
Number of parameters:	35	262143
geom. mean likelihood:	0.63614936	0.74154728

Information Criteria:

AIC-Index:	6762.55	528710.25
BIC-Index:	6903.21	1582156.33
CAIC-Index:	6938.21	1844299.33

Power Divergence GoF statistics:

emp. value	chi-square	p-value
Cressie Read:	23910.15	p= 1.0000
Pearson Chisquare:	267588.36	p= 1.0000

=====

Likelihood ratio:	2268.31	p= 1.0000
Freeman-Tukey Chi^2:	1944.69	p= 1.0000

Degrees of freedom: 262108

WARNING: Number of cells is larger than number of different patterns!!!

obs.patterns/cells = 0.001144409179687500
number of zero cells = 261844

WARNING: Number of cells is larger than number of subjects!!!
subjects/cells = 0.001567840576171880

The data might be very sparse, please do not use the chi square p-value approximation for the Power Divergence. Goodness of Fit Statistics. Consider to use the parametric bootstrap procedure instead. In addition, several start values should be used(see defaults menu) in order to examine the occurrence of local likelihood maxima.

Der Bootstrap zeigt eine Modellgeltung auf (Tabelle 40).

Tabelle 40: WINMIRA: Übersicht Ergebnisse Bootstrap - empirischer Teil I

No.:	Satlik/LogLik	LR	CressieRead	Pearson X^2	FT
Z:		3.779	1.260	0.185	3.3198
P(X>Z):		0.000	0.104	0.426	0.0005
Mean:	2025.357		16486.707	205597.152	1810.5434
Stdev:	64.294		5890.330	334444.662	40.4069
p-values					
(emp. PDF):	0.000		0.063	0.165	0.0000

Die Cressie-Read-Statistik ($p = 0.063$) und die Pearson χ^2 -Statistik ($p = 0.165$) zeigen jeweils ein nicht signifikantes Ergebnis an.

8.2.1.3 Rasch-Analyse mit eRm

Tabelle 41 gibt einen Überblick über die Ergebnisse der Rasch-Analyse mit eRm.

Tabelle 41: eRm: Übersicht Ergebnisse Rasch-Analyse - empirischer Teil I

Results of RM estimation:				
Conditional log-likelihood: -2252.985				
Number of iterations: 46				
Number of parameters: 17				
Item (Category) Difficulty Parameters (eta) with 0.95 CI:				
	Estimate	Std. Error	lower CI	upper CI
V2	-0.840	0.118	-1.071	-0.608
V3	1.062	0.156	0.757	1.367
V4	-1.443	0.118	-1.674	-1.212
V5	-0.585	0.120	-0.819	-0.350
V6	-0.067	0.126	-0.314	0.180
V7	-2.050	0.123	-2.292	-1.808
V8	-1.594	0.119	-1.827	-1.361
V9	0.544	0.139	0.271	0.816
V10	-2.262	0.127	-2.511	-2.014
V11	2.044	0.203	1.645	2.442
V12	0.425	0.136	0.159	0.692
V13	2.496	0.235	2.036	2.956
V14	0.350	0.134	0.087	0.613
V15	-0.379	0.122	-0.617	-0.141
V16	0.711	0.144	0.429	0.993
V17	2.618	0.245	2.138	3.097
V18	0.689	0.143	0.408	0.970
Item easiness Parameters (beta) with 0.95 CI:				
	Estimate	Std. Error	lower CI	upper CI
beta V1	1.719	0.120	1.485	1.954
beta V2	0.840	0.118	0.608	1.071

beta V3	-1.062	0.156	-1.367	-0.757
beta V4	1.443	0.118	1.212	1.674
beta V5	0.585	0.120	0.350	0.819
beta V6	0.067	0.126	-0.180	0.314
beta V7	2.050	0.123	1.808	2.292
beta V8	1.594	0.119	1.361	1.827
beta V9	-0.544	0.139	-0.816	-0.271
beta V10	2.262	0.127	2.014	2.511
beta V11	-2.044	0.203	-2.442	-1.645
beta V12	-0.425	0.136	-0.692	-0.159
beta V13	-2.496	0.235	-2.956	-2.036
beta V14	-0.350	0.134	-0.613	-0.087
beta V15	0.379	0.122	0.141	0.617
beta V16	-0.711	0.144	-0.993	-0.429
beta V17	-2.618	0.245	-3.097	-2.138
beta V18	-0.689	0.143	-0.970	-0.408

Die Überprüfung des Modells erfolgt mittels Andersens Likelihood-Quotienten-Test. Als Splitkriterium wird der Mittelwert der Personen-Rand-summen verwendet. Die Ergebnisse sind Tabelle 42 zu entnehmen.

Tabelle 42: eRm: Ergebnisse Modellgeltungstest - empirischer Teil I

Andersen LR-test:

LR-value: 26.657

Chi-square df: 17

p-value: 0.063

Es liegt bei einem p -Wert von 0,063 keine signifikante Modellverletzung vor. Die Schätzungen der Aufgabenparameter unterscheiden sich nicht signifikant für Personen mit niedrigen und hohen Randsummen. Aufgrund des niedrigen p -Wertes lohnt es sich, einen Blick in die Ergebnisse der Aufgabenanalyse zu werfen, um zu überprüfen, welche Aufgaben zu einem niedrigen p -Wert führen, wird der grafische Modellgeltungstest mit 95%-Konfidenz-Regionen durchgeführt. Als Splitkriterium wird der Mittelwert der Personen-Randsummen gewählt (Abbildung 13). Mit diesem Verfahren kann die Personenhomogenität überprüft werden. Dem Test liegt die Annahme zugrunde, dass dieser unabhängig von der Stichprobenzusammensetzung dieselbe Eigenschaft misst. Dies bedeutet, dass die zu überprüfenden Items stichprobenunabhängig sein müssen (vgl. Bühner 2006, S. 343).

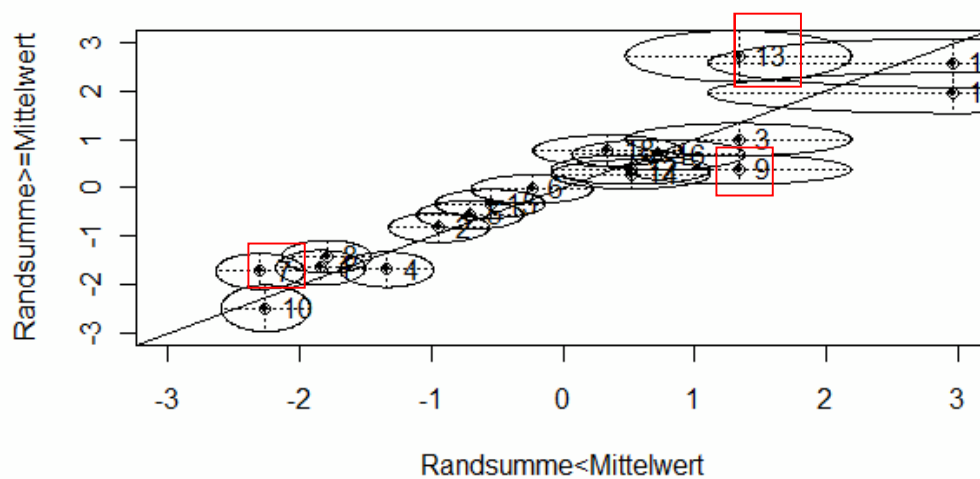


Abbildung 13: eRm: Ergebnis grafischer Modellgeltungstest - empirischer Teil I

Eine signifikante Abweichung von der Diagonalen wird bei den Aufgaben 7, 9 und 13 deutlich. Es gibt somit einen signifikanten Unterschied in der Schätzung der Aufgaben-Parameter für Personen mit niedrigen und hohen Randsummen. Mit dem aufgabenspezifischen Wald-Test lässt sich überprüfen, ob die Aufgaben 7, 9 und 13 ein signifikantes *Differential Item Functioning* darstellen (vgl. Strobel 2010). Das Splitkriterium ist wieder der Mittelwert der Personen-Randsummen (Tabelle 43).

Tabelle 43: eRm: Ergebnisse Wald-Test - empirischer Teil I

Wald test on item level (z-values):		
	z-statistic	p-value
beta V1	0.746	0.456
beta V2	0.488	0.626
beta V3	-0.711	0.477
beta V4	-1.338	0.181
beta V5	0.580	0.562
beta V6	0.747	0.455
beta V7	2.265	0.024
beta V8	1.529	0.126
beta V9	-2.056	0.040
beta V10	-0.770	0.442
beta V11	-1.029	0.303
beta V12	-0.398	0.690
beta V13	2.704	0.007
beta V14	-0.683	0.495
beta V15	0.864	0.387
beta V16	-0.106	0.916
beta V17	-0.389	0.697
beta V18	1.299	0.194

Die Parameter-Schätzungen für die Aufgaben 7, 9 und 13 zeigen bei einem p -Wert von unter 0,05 einen signifikanten Unterschied. Dies hat zur Folge, dass diese Items ausgeschlossen werden müssen. Nach dem Löschen der Aufgaben 7, 9 und 13 wird erneut Andersens Likelihood-Quotienten-Test zur Überprüfung der Modellgeltung durchgeführt. Die wiederholte Modellüberprüfung zeigt, dass erneut keine signifikante Modellverletzung vorliegt (Tabelle 44: p -Wert = 0,422). Der Wert hat sich nach der Itemreduktion von 0,063 auf 0,422 deutlich verbessert.

Tabelle 44: eRm: Ergebnisse Modellgeltungstest - empirischer Teil I

Andersen LR-test:	
LR-value:	14.373
Chi-square df:	14
p-value:	0.422

Abbildung 14 zeigt die Item-Characteristic-Curves (ICC) der 18 Items. Aufgrund der Parallelität der ICC kann angenommen werden, dass die Items die gleichen Trennschärfen aufweisen (vgl. Strobel 2010).

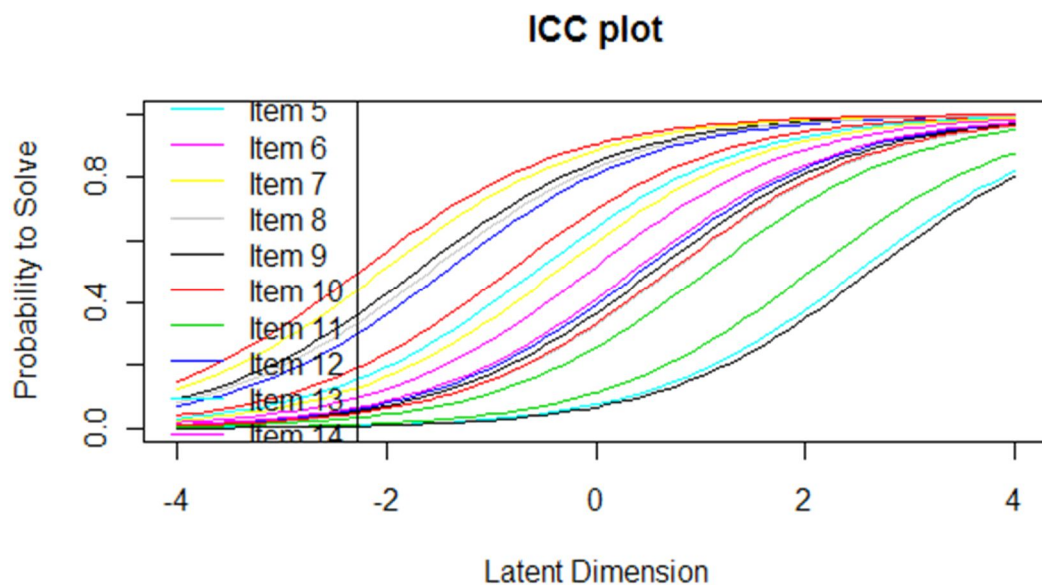


Abbildung 14: eRm: Item Characteristic Curves der 18 Items - empirischer Teil I

Mithilfe der Person-Item-Map lässt sich die Lage der Personen- und Aufgaben-Parameter auf der latenten Skala aufzeigen (Abbildung 15). Im mittlere-

ren Fähigkeits-Bereich liegen mehr Aufgaben als an den Rändern. Dementsprechend ist die Schätzung der Personen-Parameter im mittleren Bereich genauer. Dies führt zu kleineren Standardfehlern (vgl. Strobel 2010).

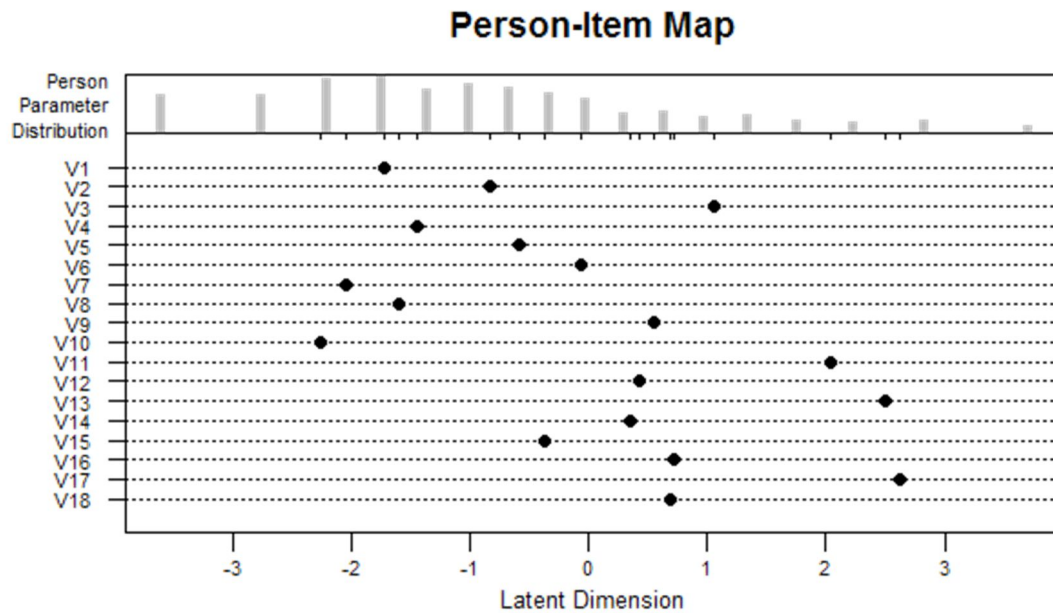


Abbildung 15: eRm: Person-Item-Map - empirischer Teil I

Tabelle 45 lässt sich entnehmen, dass jeder Proband, der neun Aufgaben richtig gelöst hat, eine geschätzte Fähigkeit von ca. 0,02 aufweist. Der Standardfehler liegt bei ca. 0,56.

Tabelle 45: eRm: Übersicht Personenparameter - empirischer Teil I

Person Parameters:			
Raw Score	Estimate	Std.Error	
0	-4.51400403	NA	
1	-3.60734314	1.0683984	
2	-2.76872878	0.8058002	
3	-2.21003290	0.7002860	
4	-1.76146394	0.6439385	
5	-1.36991958	0.6100078	
6	-1.01172818	0.5884971	
7	-0.67392383	0.5750096	
8	-0.34805774	0.5676002	
9	-0.02752768	0.5655953	
10	0.29380678	0.5690888	
11	0.62255919	0.5787532	
12	0.96668079	0.5958239	
13	1.33662049	0.6223619	
14	1.74748040	0.6621042	
15	2.22434164	0.7234186	
16	2.81999519	0.8306634	
17	3.70147386	1.0892173	
18	4.65209970	NA	

Werden die Ergebnisse aus der Person-Item-Map zum Vergleich herangezogen, kann festgestellt werden, dass die Standardfehler im mittleren Bereich am geringsten sind. Dies rührt daher, dass in diesem Bereich viele Aufgaben liegen und auch viel Information zur Schätzung der Personen-Parameter vorhanden ist (vgl. Strobel 2010).

8.3 Zusammenfassende Betrachtung des achten Kapitels

In Kapitel 8 wurde die Rasch-Analyse mit dem dichotomen Rasch-Modell (1-PL-Rasch-Modell) durchgeführt. Die Ergebnisse der Modellgeltungstests fielen unter Anwendung verschiedener Softwares sehr unterschiedlich aus. Die Rasch-Analyse mit RUMM 2030 zeigte, dass unter Verwendung des kompletten Itemsatzes des TeMaTex eine Modellverletzung vorliegt. Erst nach der Elimination der Items 2, 5, 13 und 15 lag keine Modellverletzung mehr vor. Nach den Berechnungen mit WINMIRA und eRm lagen keine Modellverletzungen vor. Innerhalb der Analysen mit WINMIRA wurde zur Überprüfung der Modellgeltung das Bootstrap-Verfahren durchgeführt, was wiederum zu keiner Modellverletzung führte, wobei hinzuzufügen ist, dass Item 13 bei allen durchgeführten Analysen kritische Werte aufzeigte und innerhalb des TeMaTex das schwierigste Item ist. Aufgrund der Tatsache, dass Item 13 in jeder Auswertungsvariante auffällig wurde, wird dieses Item für den CareMaTex nicht berücksichtigt und somit aus dem Test entfernt.

IV. Empirischer Teil 2

In Kapitel 9 werden die Ergebnisse der Expertenbefragung respektive des Expertenratings in Bezug auf die Items des CareMaTex, welche zuvor an den Kontext der Altenpflege angepasst wurden, ausführlich vorgestellt. In Kapitel 10 werden die Ergebnisse der deskriptiven und inferenzstatistischen Analysen in Bezug auf den CareMaTex vorgestellt. Die Darstellung des CareMaTex findet sich in Kapitel 11.

9 Expertenrating in Bezug auf die Items des CareMaTex

Im folgenden Kapitel 9.1 werden die Ergebnisse der Expertenbefragung vorgestellt. Hierbei wird auf das Antwortverhalten der Experten eingegangen und auf die Itemauswahl aufgrund des Expertenurteils.

9.1 Expertenbefragung

Im Folgenden wird detailliert dargestellt, wie die Experten die an den Kontext der Altenpflege angepassten Items 1 bis 17 hinsichtlich ihrer Relevanz bewertet haben. Basierend auf diesen Erkenntnissen wird beschrieben, ob weitere Items aus dem Test aufgrund einer fehlenden inhaltlichen Validität entfernt werden müssen.

9.1.1 Antwortverhalten der Experten

In Tabelle 46 wird das Antwortverhalten der Rater in Häufigkeiten ausführlich dargestellt. Die Tabelle liest sich wie folgt: Item 1 wurde von den Experten einmal als sehr relevant, dreimal als relevant, etc. beurteilt. Bei der Evaluationsfrage 18 wurde eine andere Bezeichnung für die Skalierung gewählt, nämlich von *trifft zu* bis *trifft nicht zu*.

Tabelle 46: Übersicht Antwortverhalten

Items (Evaluationsfragen)	sehr relevant	relevant	wenig relevant	nicht relevant
1	1	3	2	1
2	1	3	1	2
3	1	0	5	1
4	1	3	2	1
5	1	2	2	2
6	1	3	1	1
7	3	3	0	1
8	3	3	0	1
9	1	5	0	1
10	4	2	0	1
11	1	4	1	1
12	2	3	1	1
13	1	3	2	1
14	1	5	0	1
15	1	4	0	2
16	0	0	4	3
17	1	4	1	1
	trifft zu	trifft eher zu	trifft eher nicht zu	trifft nicht zu
18	0	3	2	2

9.1.2 Itemanalyse aufgrund des Expertenurteils

Tabelle 47 gibt die Anzahl positiver und negativer Einschätzungen bezüglich der 17 zu bewertenden Items an.

Tabelle 47: Auswertung Expertenrating

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Item 17
Anzahl positiver Einschätzungen	4	4	1	4	3	4	6	6	6	6	5	5	4	6	5	0	5
Anzahl negativer Einschätzungen	3	3	6	3	4	2	1	1	1	1	2	2	3	1	2	7	2
positive Übereinstimmung in %	57	57	14	57	43	57	86	86	86	86	71	71	57	86	71	0	71

Von den 17 Items des CareMaTex erhielten in dieser Expertenbefragung fünf Items eine Zustimmung von 86 %, vier Items eine Zustimmung von 71 %, fünf Items lagen bei 57 % und drei Items erhielten weniger als 50 % Zustimmung. Die Evaluationsfrage 18 erhielt von den Experten nur 43 % Zustimmung. Das Ergebnis zu Evaluationsfrage 18 ist überraschend, obwohl die überwiegende Anzahl der 17 Items von den Experten als *sehr relevant* oder *relevant* eingestuft wurden. Daher werden die Ergebnisse bezüglich der Evaluationsfrage 18 mit Vorsicht interpretiert und für die weiteren Schritte zur Überarbeitung der Testversion nicht unkritisch übernommen. Aufgrund der Tatsache, dass Item 3 und 16 von allen Experten als *wenig relevant* bis *nicht relevant* eingestuft wurden, werden diese beiden Items näher betrachtet. Item 3 ist eine Aufgabe zum Aufgabentext 1 *Renovierung eines Altenpflegeheimes*. Folgende Aufgabenstellung war gegeben: *Berechnen Sie, wie groß die Fläche ist, die von der Firma Renofix GmbH für das Streichen der Wände in Raum 0.06 in Rechnung gestellt wird. Beachten Sie dabei den Abschnitt (2) des Vertragsauszuges!* Tabelle 48 verdeutlicht den Lösungsweg von Aufgabe 3.

Tabelle 48: Lösungsweg zur Aufgabe 3

Aufgabe	IALS-Stufe	Mathematische Anforderungen	Rechenoperationen
3	Stufe 5	Schritt: Umfangberechnung/ Arithmetik: Rechteck/ Multiplikation	$U = 2 \cdot 15 \text{ m} + 2 \cdot 20 \text{ m}$
		Schritt: Berechnung der Mantelfläche	$M = (2 \cdot 15 \text{ m} + 2 \cdot 20 \text{ m}) \cdot 3 \text{ m}$
		Schritt: Prozentrechnung: Prozentwert berechnen	$P = \frac{210 \text{ m}^2 \times 20}{100}$
		Schritt: Arithmetik: Subtraktion	$210 \text{ m}^2 - 42 \text{ m}^2$

Im Kommentarteil zu Aufgabe 3 gaben die befragten Experten an, dass die Aufgabe aufgrund eines zu hohen Komplexitätsgrades und der vielen Teilschritte *wenig* bis *nicht relevant* eingestuft wird. Jedoch, so eine Rückmeldung, sollte die Prozentrechnung von den Auszubildenden beherrscht werden.

Eine ähnliche Begründung zeigten die Ergebnisse der Befragung hinsichtlich des schlechten Abschneidens des Aufgabentextes 16 – ehemals Aufgabe 17 im TeMaTex. Bei Aufgabe 16 handelt es sich inhaltlich um die Berechnung der Länge einer Leiter, welche benötigt wird, um in einem Materiallager an Verbandmaterial zu gelangen, welches sich 3 m über dem Boden befindet. Für ein besseres Verständnis wird im Folgenden der Lösungsweg für Aufgabe 16 vorgestellt.

Tabelle 49: Lösungsweg zur Aufgabe 16

Aufgabe	IALS-Stufe	Mathematische Anforderungen	Rechenoperationen
16	Stufe 5	Schritt: Satz des Pythagoras: Aufstellen der Formel	$l^2 = 3^2 + 4^2$
		Schritt: Auflösen der Gleichung	$l = \sqrt{3^2 + 4^2}$
		Schritt: Radizieren von Quadratzahlen	$\sqrt{25}$

Die Experten gaben an, dass sie Textverstehen und räumliche Vorstellungskraft für den Altenpflegeberuf für primär notwendig halten. Jedoch ist der Rechenweg dieser Aufgabe zu komplex.

Der Text zur Aufgabe 12 wurde ergänzt. Hier fehlte die Angabe, um welchen Cocktail es sich handelte.

9.1.3 Ermittlung der Urteilsübereinstimmung

Die Berechnung der Urteilsübereinstimmung wurde mit der Statistiksoftware R unter Verwendung des package *irr* (Various Coefficients of Interrater Reliability and Agreement) durchgeführt. Die exakten Eingabebefehle zur Berechnung des Konkordanzkoeffizienten Kendall W können Anhang 4 entnommen werden. Im Folgenden soll der Prozess aufgezeigt werden, wie der Koeffizient berechnet und interpretiert wird.

Problem: In der vorliegenden Arbeit soll überprüft werden, ob die Experten (Rater) in der Lage sind, die 17 Items der angepassten Version des TeMaTex nach ihrer Relevanz übereinstimmend zu beurteilen.

Untersuchungsaufbau: m = Sieben Experten wurden gebeten, die N = 17 überarbeiteten Items aufgrund ihrer Relevanz in Bezug auf die benötigten

mathematischen Fähigkeiten unabhängig voneinander in eine Rangreihe zu bringen (Rangplatz 1: sehr relevant; Rangplatz 6: nicht relevant). Tabelle 50 zeigt die Ergebnisse.

Tabelle 50: Ergebnisse des Expertenratings

Experte	Items																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	2	2	3	2	3	2	2	2	2	2	2	1	2	2	2	3	2
2	1	1	3	1	1	1	1	1	1	1	3	3	1	1	1	3	1
3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
4	2	2	3	2	2	2	1	1	2	1	2	1	3	2	4	4	2
5	3	3	3	3	3	0	2	2	2	1	1	2	2	2	2	3	3
6	3	4	3	3	4	3	2	2	2	2	2	2	3	2	2	4	2
7	2	2	3	2	2	2	1	1	2	1	2	2	2	2	2	3	2
Rangsummen (T _i)	17	18	22	17	19	14	13	13	15	12	16	15	17	15	17	24	16

Nullhypothese (H₀): Die Rangreihen der Experten stimmen überein.

Alternativhypothese (H₁): Die Rangreihen der Experten stimmen nicht überein.

Signifikanzniveau: Alpha = 0,05

Durchführung: Tabelle 50 ist zu entnehmen, dass zum Beispiel bei Item 1 von einem Experten auf Rangplatz 1, von drei Experten auf Rangplatz 2, von zwei Experten auf Rangplatz 3 und von einem Experten auf Rangplatz 4 gesetzt wurde. Tabelle 51 stellt die Ergebnisse der Berechnungen mit dem Statistikprogramm R unter Anwendung des package *irr* für die gesamte Matrix dar.

Tabelle 51: Übersicht Ergebnisse der Berechnung Urteilsübereinstimmung

```
kendall's coefficient of concordance w
Subjects = 17
Raters = 7
w = 0.257
chisq(16) = 28.8
p-value = 0.0254
```

Entscheidung: Aufgrund der schwachen Urteilsübereinstimmung ($W = 0,257$) und des signifikanten p-Wertes von 0,0254 ist H_0 zugunsten H_1 zu verwerfen.

9.2 Zusammenfassende Betrachtung des neunten Kapitels

In Kapitel 9 wurde die Durchführung des Expertenratings beschrieben. Es sei festzuhalten, dass es keine Übereinstimmung der Experten gab. Daher verbleiben vorläufig alle 17 Items im CareMaTex, obwohl bei Aufgabe 16 die geforderte mathematische Operation laut der befragten Experten nicht als notwendige mathematische Anforderung für den Beruf der Altenpflege gesehen wird. Dies bestätigten bereits die Analysen in Kapitel 3.3.

10 Statistische Analysen in Bezug auf den CareMaTex

Im folgenden Kapitel 10.1 werden die Ergebnisse der deskriptiven Analyse vorgestellt. Hierbei wird auf die Zusammensetzung der Stichprobe eingegangen sowie auf die Antworthäufigkeiten der Probanden, unterteilt nach Schulabschlüssen. Die Darstellung der Ergebnisse der inferenzstatistischen Analyse findet in Kapitel 10.2 statt. In diesem Kapitel wird die Modelentwicklung mithilfe der verschiedenen Statistikprogramme erläutert.

10.1 Beschreibende statistische Analyse

Die Stichprobe umfasste insgesamt 253 Probanden. Im Rahmen der Datenerhebung wurden Angaben zum Schulabschluss der Probanden erhoben. Tabelle 52 gibt einen Überblick über die Verteilung der Schulabschlüsse in der Stichprobe.

Tabelle 52: Verteilung Schulabschlüsse - empirischer Teil II

Hauptschulabschluss	Realschulabschluss	Fachhochschulreife/Abitur	Keine Angaben
94	107	38	14

Der Realschulabschluss stellt mit 42,3 % den am häufigsten angegebenen Schulabschluss in der Stichprobe dar, gefolgt vom Hauptschulabschluss mit 37,2 %. Probanden mit Fachhochschulreife beziehungsweise mit Abitur machen hingegen nur 15,0 % aus. Circa 5,5 % machten keine Angaben zu ihrem Schulabschluss.

Im Folgenden werden die im Expertenrating verwendeten 17 Items in Bezug auf ihre Antworthäufigkeiten beschrieben. In der zweiten Spalte von Tabelle 53 werden die Items und in der letzten Spalte die Anzahl der getesteten Probanden angegeben. Die Spalten 4 und 5 geben die Häufigkeiten der gegebenen Antworten zu den jeweiligen Items an (0 = keine oder falsche Antwort, 1 = richtige Antwort). Aus der Tabelle geht hervor, wie häufig eine Aufgabe von den Testpersonen gelöst oder nicht gelöst beziehungsweise nicht bearbeitet wurde. Item 12 (Aufgabe 12 des Aufgabentext 3: Diabetes und Alkohol) wurde von den Probanden am häufigsten nicht gelöst oder

nicht beantwortet (0 = 242). Die Probanden beantworteten dieses Item lediglich 11-mal richtig. Item 7 (Aufgabe 7 des Aufgabentextes 2: Der ambulante Pflegedienst) wurde am häufigsten richtig gelöst (1 = 149).

Tabelle 53: Antworthäufigkeiten der 17 Items des CareMaTex – empirischer Teil II

Item labels and sample frequencies:					
no.	n of label	categories	0	1	N
1	VAR1	2	118	135	253
2	VAR2	2	172	81	253
3	VAR3	2	231	22	253
4	VAR4	2	130	123	253
5	VAR5	2	185	68	253
6	VAR6	2	198	55	253
7	VAR7	2	104	149	253
8	VAR8	2	175	78	253
9	VAR9	2	230	23	253
10	VAR10	2	108	145	253
11	VAR11	2	227	26	253
12	VAR12	2	242	11	253
13	VAR13	2	229	24	253
14	VAR14	2	170	83	253
15	VAR15	2	213	40	253
16	VAR16	2	239	14	253
17	VAR17	2	220	33	253

Im weiteren Verlauf wurden die Häufigkeiten der empirisch erreichten Gesamtpunktzahlen nach den unterschiedlichen Schulabschlüssen aufgeteilt. Aus der Tabelle 54 können die Häufigkeiten der Gesamtpunktzahlen entnommen werden. In der linken Spalte der Tabelle stehen die theoretisch erreichbaren Gesamtpunktzahlen (0 bis 17 Punkte). In der zweiten, dritten und vierten Spalte werden die Häufigkeiten der jeweils erreichten Gesamtpunktzahlen genannt.

Tabelle 54: Verteilung der Gesamtpunktwerte nach Schulabschluss – empirischer Teil II

Ge- samt- punkt- werte	Hauptschul- abschluss	Realschul- abschluss	Fachhochschul- abschluss/Abitur
0	13	13	1
1	13	12	3
2	16	10	4
3	14	13	0
4	8	10	4
5	9	8	5
6	7	6	2
7	7	15	2
8	2	8	1
9	3	5	3
10	2	4	3
11	0	0	2
12	0	1	2
13	0	0	3
14	0	0	1
15	0	1	1
16	0	1	1
17	0	0	0
Gesamt	94	107	38

Abbildung 16 zeigt einen Vergleich zwischen den Personenlokalisationen und den Itemlokalisationen. Die Logitskala umfasst Werte im Bereich von -5 und 4. Die durchschnittlichen Personenfähigkeiten liegen mit einem Wert von -1,715 unterhalb der durchschnittlichen Itemschwierigkeiten von 0,000. Bei circa 92 Personen ist die Personenfähigkeit niedriger als die geringste Itemschwierigkeit. Dies lässt vermuten, dass die Items des Tests gegebenenfalls zu schwer für die Personen sind und dass die Items für eine Differenzierung innerhalb der untersuchten Personengruppe nicht geeignet sind.

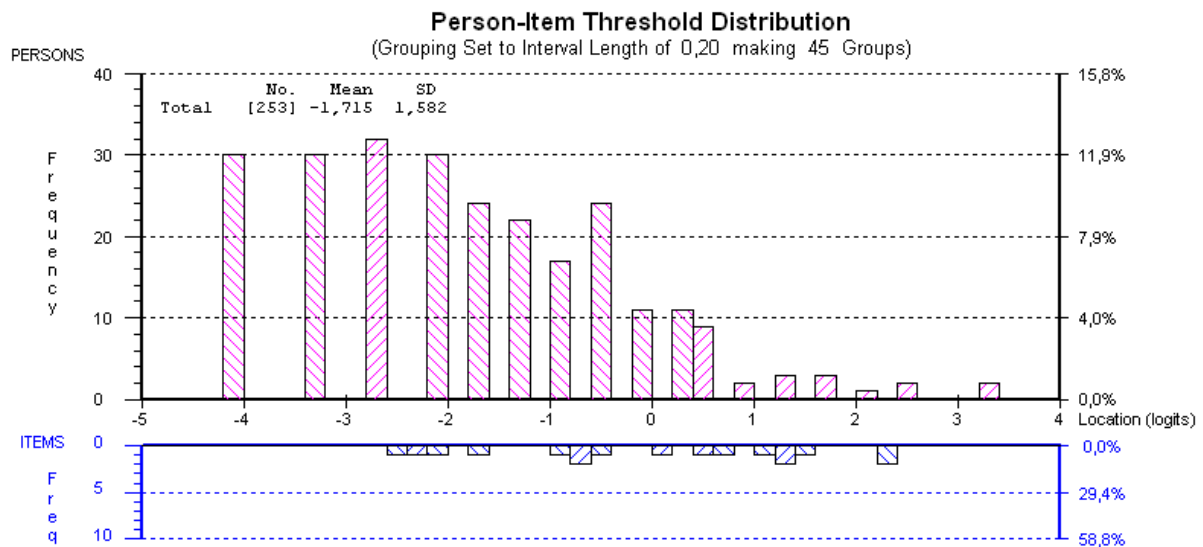


Abbildung 16: RUMM 2030: Verteilung Personenfähigkeit und Itemschwierigkeit – empirischer Teil II

10.2 Schließende statistische Analyse

Mithilfe der Parameterschätzung im dichotomen Rasch-Modell erlangen die Daten bei Modellpassung das entsprechende Skalenniveau. Beim Vorliegen der spezifischen Objektivität sind die Ergebnisse stichprobenunabhängig (vgl. Moosbrugger 2007d, S. 230 f.).

10.2.1 Ergebnisse aus der Modellentwicklung anhand des Rasch-Modells

Um die Konstruktvalidität des CareMaTex ermitteln zu können, wurde die Rasch-Analyse durchgeführt. In die Rasch-Analyse wurden insgesamt $n = 253$ Probanden einbezogen (Tabelle 55).

Tabelle 55: Allgemeine Daten zur Rasch-Analyse - empirischer Teil II

number of persons	:	253
number of items	:	17
number of classes	:	1
max. number of iterations	:	250
accuracy criterion	:	0.0005
random start value	:	4321

Zur Analyse der Daten wurden erneut die Statistikprogramme RUMM 2030, WINMIRA und in eRm eingesetzt. Die Auswertungen der verschiedenen

Softwareprogramme befinden sich auf der Daten-CD unter den entsprechenden Ordnern. Zu Beginn werden die Daten mit RUMM 2030 ausgewertet und anschließend mit den Ergebnissen der Analysen aus WINMIRA und eRm abgeglichen. Der CareMaTex umfasst 17 Items (Tabelle 55). Daraus lassen sich 131.072 mögliche theoretische Antwortmuster erstellen (Tabelle 56). Empirisch liegen allerdings nur 167 Antwortmuster vor, sodass erneut die Voraussetzungen für eine χ^2 -Asymptotik des globalen Modellgeltungstests nicht vorliegen. Hierfür müsste jedes Antwortmuster mindestens einmal auftreten (vgl. Rost 2004).

Tabelle 56: Überblick Rasch-Analyse – empirischer Teil II

saturated likelihood:	-1181.0766
number of different patterns:	167
number of possible patterns:	131072

10.2.1.1 Prüfung der Modellgeltung mit RUMM 2030

Zunächst werden Angaben zur Teststruktur, zur Stichprobe und zu den Parameterschätzungen gemacht (Tabelle 57).

Tabelle 57: RUMM 2030: Überblick Rasch-Analyse - empirischer Teil II

```

** TEST STRUCTURE
No. of Items      17
No. of Categories  Equal across Items : set at 2 categories
per item
Score Range [All Items]  16

** CALIBRATING SAMPLE
No. of Persons:
* entered Project   :    253
* invalid records   :      0
* extreme scores    :    30
* valid scores      :   223 [available for analysis]
Missing data detected: None

** ESTIMATION DETAILS
Item Converge Limit 0,00010
Iterations [Items]   5 [all 17 parameters converged]
Iterations [Persons] -1
Lowest score         1
Highest score        16
Person Converge Limit 0,01000
Person Estimation    Weighted Maximum Likelihood method
Extreme Pers. Criterion 0,220

```

Zum Zeitpunkt der Analyse liegen keine ungültigen oder fehlenden Werte vor. Insgesamt wurden daher 223 gültige Fälle berücksichtigt (Tabelle 57). Für die Schätzungen der Itemparameter wurden fünf Iterationen (Wiederholungen) benötigt. Im Folgenden werden die ICC des Items 14 und 16 dargestellt.

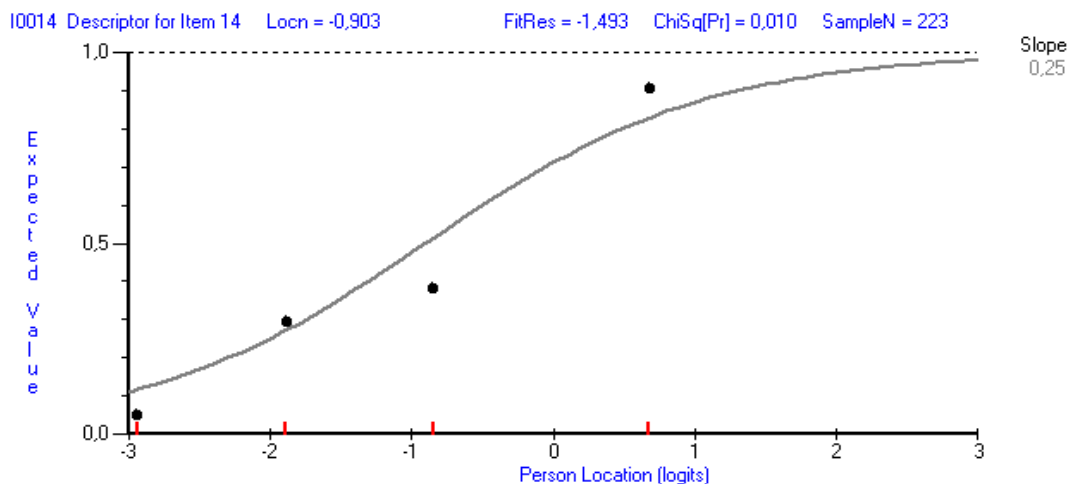


Abbildung 17: RUMM 2030: Item Characteristic Curve von Item 14 - empirischer Teil II

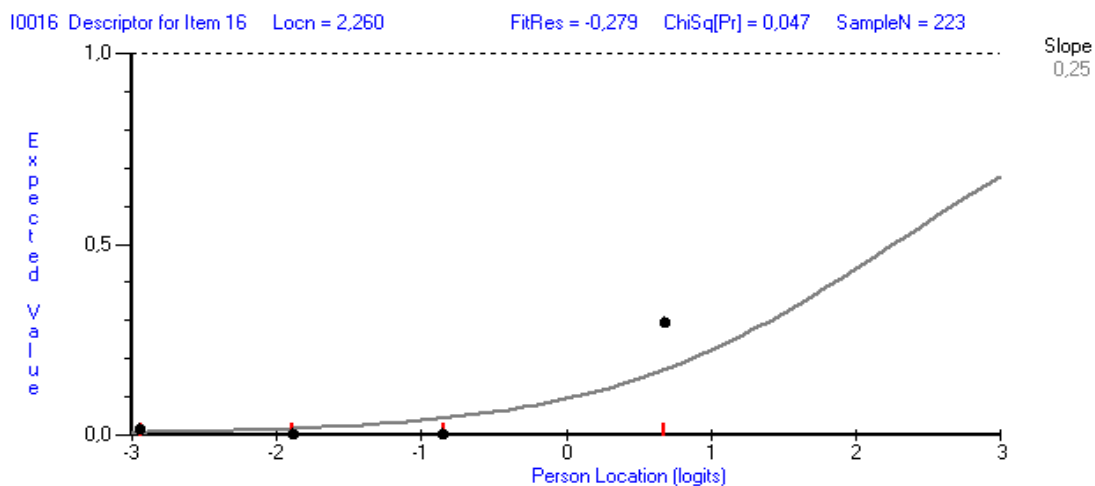


Abbildung 18: RUMM 2030: Item Characteristic Curve von Item 16 - empirischer Teil II

Bei der Betrachtung der einzelnen ICC der 17 Items fällt auf, dass Item 7 mit einer Item-Schwierigkeit von -2,405 das leichteste zu lösende Item darstellt und Item 16 mit einer Item-Schwierigkeit von 2,260 das schwierigste.

Tabelle 58: RUMM 2030: Übersicht Item-Person und Item-Trait Interaction - empirischer Teil II

ITEM-PERSON INTERACTION						
	ITEMS			PERSONS		
	Location	Fit	Residual	Location	Fit	Residual
Mean	0,000	-0,424		-1,715	-0,289	
SD	1,530	1,001		1,582	0,702	
Skewness		0,741			0,577	
Kurtosis		-0,599			0,164	
Correlation		0,000			0,241	
Complete data DF =		0,937				
ITEM-TRAIT INTERACTION				RELIABILITY INDICES		
Total Item Chi Squ		71,146		Separation Index	0,73561	
Total Deg of Freedom		51,000		Cronbach Alpha	0,83082	
Total Chi Squ Prob		0,032597				
LIKELIHOOD-RATIO TEST			POWER OF TEST-OF-FIT			
Chi Squ			Power is GOOD			
Degrees of Freedom			[Based on SepIndex of 0,73561]			
Probability						

Im Rahmen der Modellentwicklung werden die Fit-Residuen der 17 Items genauer betrachtet. Aus der Tabelle 59 lässt sich entnehmen, dass keines der 17 Items über dem kritischen Wert liegt. Im Rahmen der Modellüberprüfung wurden die χ^2 -Werte der einzelnen Items betrachtet. Item 14 und Item 16 fielen hierbei mit einem signifikanten p-Wert auf (Tabelle 59).

Tabelle 59: RUMM 2030: Übersicht Individual Item-Fit Werte - empirischer Teil II

Seq	Item	Type	Location	SE	Residual	DF	ChiSq	DF	Prob
1	I0001	Poly	-2,022	0,158	1,768	208,94	7,426	3	0,059486
2	I0002	Poly	-0,747	0,162	-1,059	208,94	2,718	3	0,437136
3	I0003	Poly	1,477	0,251	-1,359	208,94	1,262	3	0,738251
4	I0004	Poly	-1,713	0,156	1,190	208,94	1,676	3	0,642230
5	I0005	Poly	-0,410	0,168	-0,815	208,94	2,087	3	0,554561
6	I0006	Poly	0,052	0,180	0,479	208,94	3,935	3	0,268549
7	I0007	Poly	-2,405	0,162	-0,376	208,94	6,221	3	0,101322
8	I0008	Poly	-0,623	0,164	-0,911	208,94	5,047	3	0,168362
9	I0009	Poly	1,383	0,245	-1,451	208,94	3,158	3	0,367862
10	I0010	Poly	-2,210	0,160	1,010	208,94	1,880	3	0,597745
11	I0011	Poly	1,174	0,231	-0,259	208,94	0,982	3	0,805650
12	I0012	Poly	2,234	0,318	-0,545	208,94	4,424	3	0,219146
13	I0013	Poly	1,203	0,233	-0,637	208,94	1,798	3	0,615463

14	I0014	Poly	-0,903	0,160	-1,493	208,94	11,287	3	0,010271
15	I0015	Poly	0,501	0,196	-1,721	208,94	6,449	3	0,091716
16	I0016	Poly	2,260	0,321	-0,279	208,94	7,936	3	0,047352
17	I0017	Poly	0,750	0,207	-0,750	208,94	2,859	3	0,413837

Die Ergebnisse der wiederholten Modellprüfung zeigen eine Modellgeltung für die auf den χ^2 -Werten basierende Itemreduktion an (Tabelle 60).

Tabelle 60: RUMM 2030: Ergebnisse χ^2 basierende Itemreduktion - empirischer Teil II

ITEM-TRAIT INTERACTION		RELIABILITY INDICES	

--			
Total Item Chi Squ	48,691	Separation Index	0,70419
Total Deg of Freedom	45,000	Cronbach Alpha	0,80606
Total Chi Squ Prob	0,326785		

Es wurden insgesamt zwei Items eliminiert, und zwar Item 14 und 16.

10.2.1.2 Prüfung der Modellgeltung mit WINMIRA

In der Spalte „Expected freq.“ wird die theoretisch erwartete Punkteverteilung der 253 Probanden unter Voraussetzung der Gültigkeit des Rasch-Modells dargestellt (Tabelle 61). Theoretisch müssten 30 Probanden keine und ein Proband alle Testaufgaben richtig beantwortet haben. In den nächsten Spalten sind die Schätzungen der Personenparameter durch die Maximum-Likelihood-Schätzung (MLE) und die Warm-Likelihood-Schätzung (WLE) tabellarisch aufgelistet. Die Personenparameter anhand der WLE liegen zwischen $\theta_0 = -4.554$ bei einer Gesamtpunktzahl von 0 Punkten und $\theta_{17} = 4.454$ bei einer Gesamtpunktzahl von 17 Punkten.

Tabelle 61: WINMIRA: Übersicht Personenparameter - empirischer Teil II

Raw-score	Expected freq.	MLE-estimate	std. error MLE	WLE-estimate	std. error WLE
0	30.00	*****	*****	-4.554	1.568
1	30.00	-3.629	1.081	-3.292	0.963
2	32.00	-2.761	0.825	-2.605	0.793
3	30.00	-2.171	0.723	-2.080	0.712
4	24.00	-1.688	0.670	-1.630	0.665
5	22.00	-1.263	0.637	-1.224	0.635
6	17.00	-0.871	0.616	-0.845	0.615
7	24.00	-0.501	0.603	-0.484	0.602
8	11.00	-0.143	0.595	-0.134	0.595
9	11.00	0.208	0.592	0.209	0.592
10	9.00	0.559	0.594	0.551	0.594

11		2.00		0.916		0.602		0.897		0.602
12		3.00		1.289		0.619		1.253		0.617
13		3.00		1.690		0.650		1.632		0.645
14		1.00		2.145		0.705		2.051		0.691
15		2.00		2.710		0.809		2.546		0.774
16		2.00		3.554		1.071		3.209		0.946
17		0.00		*****		*****		4.454		1.552

Der Mittelwert der Personenparameterschätzung liegt bei -1.752 (Tabelle 62). Der Anova-Reliabilitätskoeffizient (0.784) und der Reliabilitätskoeffizient nach Andrich (0.725) geben Hinweise auf eine annähernd ausreichend vorhandene innere Konsistenz.

Tabelle 62: WINMIRA: Übersicht Mittelwert Personenparameterschätzung - empirischer Teil II

WLE estimates :	Mean =	-1.752
	Var =	2.717
	stdev =	1.648
marginal error variance =		0.747
	stdev =	0.864
anova reliability =		0.784
Andrichs reliability =		0.725

Die geschätzten Itemparameter für die 17 Testaufgaben liegen im Bereich von $\theta_7 = -2.35$ und $\theta_{12} = 2.34$. Vier Items sind in den Bereich der leichteren Items unter -1 und sechs Items in den Bereich der schwereren Items über 1 anzusiedeln. Insgesamt sieben Items sind im mittleren Schwierigkeitsbereich von -1 bis 1 anzuordnen (Tabelle 63).

Tabelle 63: WINMIRA: Übersicht Itemschwierigkeiten der Items - empirischer Teil II

item label	item location	threshold parameters
VAR1	-2.02076	
VAR2	-0.73512	
VAR3	1.41193	
VAR4	-1.74255	
VAR5	-0.38747	
VAR6	-0.00215	
VAR7	-2.35056	
VAR8	-0.65736	
VAR9	1.34908	
VAR10	-2.25535	
VAR11	1.17348	
VAR12	2.34060	
VAR13	1.28851	

VAR14		-0.78627
VAR15		0.52432
VAR16		2.02842
VAR17		0.82124

Zur Bestimmung der lokalen Modellgeltung der 17 Items werden die Q-Indizes und die ZQ-Werte der jeweiligen Items analysiert (Tabelle 64).

Tabelle 64: WINMIRA: Übersicht Q-Indizes und ZQ-Werte - empirischer Teil II

itemlabel	Q-index	Zq	p (X>Zq)	
VAR1	0.1577	0.8966	0.18497	-Q...!....+
VAR2	0.1090	-0.5447	0.70702	-....!.Q..+
VAR3	0.1076	-1.0098	0.84370	-....!.Q..+
VAR4	0.1563	0.8803	0.18935	-Q...!....+
VAR5	0.1305	-0.0064	0.50255	-....Q....+
VAR6	0.1658	1.0855	0.13884	-Q...!....+
VAR7	0.0996	-0.2733	0.60770	-....!Q...+
VAR8	0.1177	-0.2989	0.61749	-....!Q...+
VAR9	0.0995	-1.1557	0.87611	-....!.Q..+
VAR10	0.1388	0.5561	0.28908	-.Q..!....+
VAR11	0.1062	-0.9079	0.81804	-....!.Q..+
VAR12	0.2191	0.4449	0.32818	-.Q..!....+
VAR13	0.1959	1.1664	0.12172	-Q...!....+
VAR14	0.1081	-0.5783	0.71847	-....!.Q..+
VAR15	0.0923	-1.2070	0.88628	-....!.Q..+
VAR16	0.0635	-1.9549	0.97471+?	-....!.Q..+
VAR17	0.1624	0.5973	0.27514	-.Q..!....+

Aus der Tabelle geht hervor, dass keiner der p-Werte signifikant ist. Als Verfahren zur Überprüfung der globalen Modellgeltung werden der Likelihood-Ratio-Test und der χ^2 -Test eingesetzt. Der Cressie-Read-Wert und der Pearson- χ^2 -Wert liegen bei $p = 1.0$, was auf eine Modellgeltung hindeutet (Tabelle 65).

Tabelle 65: WINMIRA: Übersicht Ergebnisse Modellgeltungstest - empirischer Teil II

Goodness of fit statistics:			
		estimated model	saturated model
Log-Likelihood	:	-1752.24	-1181.08
Number of parameters	:	33	131071
geom. mean likelihood	:	0.66537584	0.75987212
Information Criteria:			
AIC-Index	:	3570.48	264504.15
BIC-Index	:	3687.08	727629.05
CAIC-Index	:	3720.08	858700.05

Power Divergence GoF statistics:

		emp. value	chi-square p-value
Cressie Read	:	10279.53	p= 1.0000
Pearson Chisquare	:	88234.56	p= 1.0000

=====

Likelihood ratio	:	1142.33	p= 1.0000
Freeman-Tukey Chi^2	:	998.60	p= 1.0000

Degrees of freedom	:	131038
--------------------	---	--------

WARNING: Number of cells is larger than number of different patterns!!!
 obs.patterns/cells = 0.001274108886718750
 number of zero cells = 130905

WARNING: Number of cells is larger than number of subjects!!!
 subjects/cells = 0.001930236816406250

The data might be very sparse, please do not use the
 chi square p-value approximation for the Power Divergence
 Goodness of Fit Statistics.
 Consider to use the parametric bootstrap procedure instead.
 In addition, several start values should be used
 (see defaults menu) in order to examine the occurrence
 of local likelihood maxima.

Der Bootstrap zeigt eine Modellgeltung auf (Tabelle 66).

Tabelle 66: WINMIRA: Übersicht Ergebnisse Bootstrap - empirischer Teil II

Parametric Bootstrap estimates for Goodness of Fit:

No.:	Satlik LogLik	LR	CressieRead	Pearson X^2	FT
	Z: 3.492	0.144	-0.062	3.1518	
	P(X>Z): 0.000	0.443	0.525	0.0008	
	Mean: 986.866	8751.541	203836.813	902.1818	
	Stdev: 44.520	10622.673	1877720.089	30.5929	
p-values (emp. PDF):	0.000	0.193	0.303	0.0000	

Im Rahmen der Analysen mit RUMM 2030 mussten Item 14 und 16 eliminiert werden, um eine Modellgeltung aufzuzeigen. Um eine Vergleichbarkeit hinsichtlich der Ergebnisse der Modellgeltungstests aus RUMM 2030 und WINMIRA zu gewährleisten, werden Item 14 und 16 nachträglich aus der Itembatterie entfernt und es wird erneut ein Modellgeltungstest mit WINMIRA durchgeführt. Es konnte wiederum eine Modellgeltung festgestellt

werden. Der Cressie-Read-Wert liegt bei 0.165 und der Pearson- χ^2 -Wert bei 0.345.

10.2.1.3 Prüfung der Modellgeltung mit eRm

Im Folgenden sollen die Itemschwierigkeiten der 17 Items aufgezeigt werden (Tabelle 67).

Tabelle 67: eRm: Übersicht Itemschwierigkeiten - empirischer Teil II

beta v1	beta v2	beta v3	beta v4	beta v5	beta v6
2.020644922	0.735039034	-1.411802020	1.742424253	0.387437016	0.002117174
beta v7	beta v8	beta v9	beta v10	beta v11	beta v12
2.350432074	0.657294804	-1.348989652	2.255233916	-1.173430806	-2.340389127
beta v13	beta v14	beta v15	beta v16	beta v17	
-1.288440608	0.786186519	-0.524307634	-2.028268405	-0.821181463	

Die Itemschwierigkeiten bewegen sich in einem Bereich zwischen -2.34 (Item 12) und 2.35 (Item 7). Die Überprüfung des Modells erfolgt mithilfe von Andersens Likelihood-Quotienten-Test. Als Splitkriterium wird der Mittelwert der Personen-Randsummen verwendet. Die Ergebnisse sind Tabelle 68 zu entnehmen.

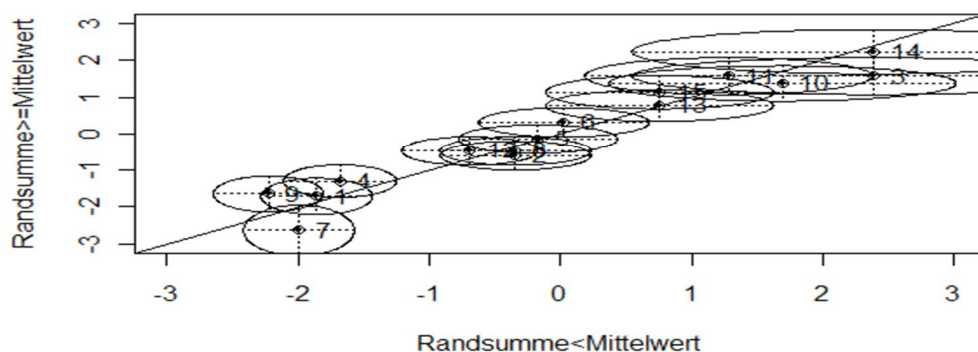
Tabelle 68: eRm: Ergebnisse Modellgeltungstest - empirischer Teil II

```
Warnung in LRtest.Rm(rm1, splitcr = "mean", se = TRUE) :
The following items were excluded due to inappropriate response patterns within subgroups:
v9 v12
Full and subgroup models are estimated without these items!
> lr1

Andersen LR-test:
LR-value: 10.584
Chi-square df: 14
p-value: 0.718
```

Im Rahmen der Überprüfung der Modellgeltung wurden Item 9 und 12 aus den Berechnungen exkludiert. Dies ist darauf zurückzuführen, „dass sich durch die Teilung der Stichproben die Personenzahl halbiert und es vor- kommen kann, dass Antwortkategorien nicht besetzt sind. Dies führt zu Problemen bei der Schätzung der Itemparameter. Daher ist die Durchfüh- rung dieses Tests manchmal nicht möglich oder nur mit einer reduzierten Itemanzahl. Das bedeutet der Test ist nur für die Items sinnvoll anwendbar, bei denen alle Antwortkategorien besetzt sind“ (Bühner 2011, S. 532). Es liegt bei einem p -Wert von 0,718 keine signifikante Modellverletzung vor. Die Schätzungen der Aufgabenparameter unterscheiden sich nicht signifi- kant für Personen mit niedrigen und hohen Randsummen. Zur Überprüfung der Personenhomogenität wurde der grafische Modellgeltungstest durchge- führt (Abbildung 19).

Abbildung 19: eRm: Ergebnis grafischer Modellgeltungstest - empirischer Teil II



Bei den Aufgaben ist keine signifikante Abweichung von der Diagonale er- kennbar. Es gibt somit keinen signifikanten Unterschied in der Schätzung der Aufgaben-Parameter für Personen mit niedrigen und hohen Randsum- men. Mit dem aufgabenspezifischen Wald-Test lässt sich überprüfen, ob eine Aufgabe ein signifikantes *Differential Item Functioning* darstellt (vgl. Strobel 2010). Das Splitkriterium ist wieder der Mittelwert der Personen- Randsummen (Tabelle 69). Auch für den Wald-Test wurden Item 9 und 12 exkludiert. Die Parameter-Schätzungen für die verbliebenen 15 Aufgaben zeigen erneut keinen signifikanten Unterschied.

Tabelle 69: eRm: Ergebnisse Wald-Test - empirischer Teil II

Warnung in waldtest.Rm(rm1, splitcr = "mean") :			
The following items were excluded due to inappropriate response patterns within subgroups:			
v9 v12			
Subgroup models are estimated without these items!			
wald test on item level (z-values):			
	z-statistic	p-value	
beta v1	0.439	0.661	
beta v2	-0.719	0.472	
beta v3	-0.840	0.401	
beta v4	1.223	0.221	
beta v5	0.074	0.941	
beta v6	0.738	0.460	
beta v7	-1.559	0.119	
beta v8	-0.368	0.713	
beta v10	1.773	0.076	
beta v11	-0.445	0.656	
beta v13	0.474	0.635	
beta v14	0.651	0.515	
beta v15	0.027	0.979	
beta v16	-0.160	0.873	
beta v17	0.759	0.448	

Abbildung 20 zeigt die Item-Characteristic-Curves (ICC) der 17 Items. Aufgrund der Parallelität der ICC kann angenommen werden, dass die Items die gleichen Trennschärfen aufweisen (vgl. Strobel 2010).

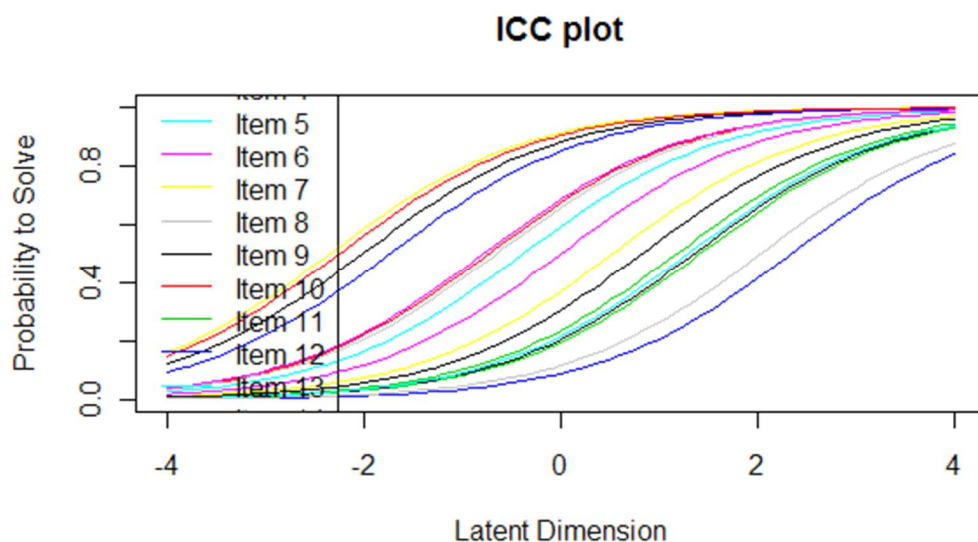


Abbildung 20: eRm: Item Characteristic Curves der 17 Items - empirischer Teil II

Mithilfe der Person-Item-Map lässt sich die Lage der Personen- und Aufgaben-Parameter auf der latenten Skala aufzeigen (Abbildung 21). Im mittleren Fähigkeits-Bereich liegen mehr Aufgaben als an den Rändern.

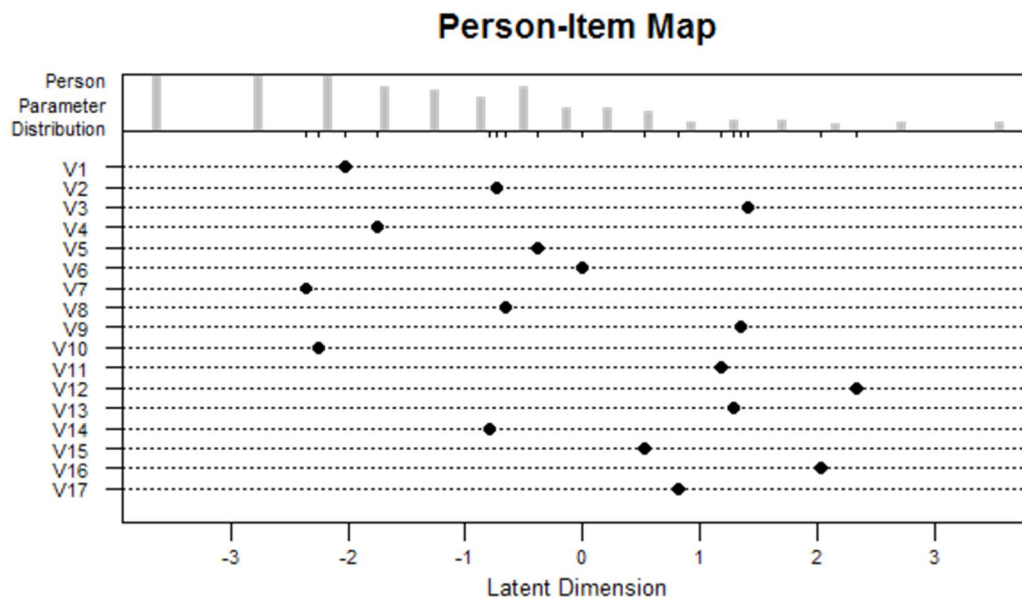


Abbildung 21: eRm: Person-Item-Map - empirischer Teil II

Dementsprechend ist die Schätzung der Personen-Parameter im mittleren Bereich genauer. Dies führt zu kleineren Standardfehlern (vgl. Strobel 2010). Aus der Tabelle 70 lässt sich entnehmen, dass jeder Proband, der neun Aufgaben richtig gelöst hat, eine geschätzte Fähigkeit von ca. 0,20 aufweist. Der Standardfehler liegt bei ca. 0,59.

Tabelle 70: eRm: Übersicht Personenparameter - empirischer Teil II

Person Parameters:		
Raw Score	Estimate	Std.Error
0	-4.5636260	NA
1	-3.6286950	1.0810026
2	-2.7611223	0.8246157
3	-2.1704922	0.7234473
4	-1.6883690	0.6697291
5	-1.2629565	0.6370788
6	-0.8712822	0.6160063
7	-0.5007437	0.6024866
8	-0.1429571	0.5946461
9	0.2084314	0.5917345
10	0.5593311	0.5939046
11	0.9163639	0.6023129
12	1.2884732	0.6194686
13	1.6897535	0.6502066
14	2.1452909	0.7046716
15	2.7097091	0.8093618
16	3.5541924	1.0715341
17	4.4668136	NA

10.3 Zusammenfassende Betrachtung des zehnten Kapitels

In Kapitel 10 wurde der CareMaTex in Bezug auf eine Modellpassung überprüft. Die Ergebnisse der Modellgeltungstests fielen unter Verwendung der verschiedenen Softwares erneut sehr unterschiedlich aus. Die Rasch-Analyse mit RUMM 2030 zeigte, dass unter Verwendung des kompletten Itemsatzes des CareMaTex eine Modellverletzung vorliegt. Erst nach der Elimination von Item 14 und 16 lag keine Modellverletzung mehr vor. Nach den Berechnungen mit WINMIRA lag keine Modellverletzung vor. Innerhalb der Analysen mit WINMIRA wurde zur Überprüfung der Modellgeltung das Bootstrap-Verfahren durchgeführt, was wiederum zu keiner Modellverletzung führte. Die Rasch-Analyse mit eRm ergab, dass erst nach der Elimination von Item 9 und 12 eine Modellgeltung zustande kam. Mittels Andersen-Test und dem grafischen Modelltest konnte eine Personenhomogenität nachgewiesen werden. Eine Itemhomogenität konnte nach der Elimination der Items 9 und 12 mittels Wald-Test ermittelt werden.

11 Beschreibung des CareMaTex

In Kapitel 11.1 wird die vorläufige Version des CareMaTex beschrieben, indem der Aufbau des Tests dargestellt wird. In Kapitel 11.2 erfolgt eine Erläuterung zu den Gütekriterien Objektivität, Reliabilität und Validität des Tests. Kapitel 11.3 stellt eine Zusammenfassung des gesamten Kapitels dar.

11.1 Darlegung der Items des CareMaTex

Im Folgenden werden die 15 Items des CareMaTex beschrieben.

Text 1: Renovierung eines Altenpflegeheimes

Für die Renovierung des Altenpflegeheims „St. Josef“ ist die Firma Renofix GmbH zuständig. Im Jahr 2015 standen im Rahmen der Umsetzung der Heimmindestbauverordnung einige Renovierungsarbeiten an. Der Einrichtungsträger Caritas hat mit der Firma Renofix GmbH einen Vertrag abgeschlossen, der die Kosten bzgl. der Renovierungsarbeiten regelt (vgl. Vertragsauszug unten).

Bearbeiten Sie die Aufgaben zu diesem Text. Beachten Sie dazu den folgenden Vertragsauszug und die Tabelle auf der folgenden Seite.

Vertragsauszug zwischen der Firma Renofix GmbH und des Einrichtungsträgers Caritas:

§ 5: Kosten für Renovierungsarbeiten:

1. Die Bezahlung für durchgeführte Renovierungsarbeiten richtet sich nicht nach der Arbeitszeit, sondern wird anhand des Materialverbrauchs berechnet. Demnach stellt die Firma Renofix GmbH dem Einrichtungsträger folgende Beträge in Rechnung: pro m² rutschfester Teppich: 15 €; pro m verlegter Bodenleisten: 6,50 €; pro m² gestrichener Wand: 1,30 €. In diesen Beträgen sind die Arbeitskosten bereits enthalten.
2. Die Firma Renofix GmbH erstellt die Rechnungen für Renovierungsarbeiten auf Grundlage der Tabelle „Raum- und Inventarplan des Altenpflegeheims St. Josef“. Bei den zu streichenden Wandflächen werden Türen und Fensterflächen nicht einzeln berechnet, sondern es werden pauschal 20 % der gesamten Wandfläche abgezogen. Die Decke wird nicht gestrichen.

Abbildung 22: Aufgabentext 1: Renovierung eines Altenpflegeheimes

Tabelle 71: Raum- und Inventarplan des Altenpflegeheims St. Josef (Erdgeschoss und 1. Obergeschoss)

	Erdgeschoss										1. Obergeschoss									
	Raum 0.01	Raum 0.02	Raum 0.03	Raum 0.04	Raum 0.05	Raum 0.06	Raum 0.07	Raum 0.08	Raum 0.09	Raum 0.10	Raum 1.01	Raum 1.02	Raum 1.03	Raum 1.04	Raum 1.05	Raum 1.06	Raum 1.07	Raum 1.08	Raum 1.09	Raum 1.10
Angaben zum Raum (alle Räume eine rechteckige Grundfläche)																				
Länge (in m)	8	6,5	7	4,7	12	15	6	8	5	4	7	4	5	20	5	3	6	12	5	7
Breite (in m)	6	4	5,5	5	12	20	7	4	3	6	5	9	6	15	5	6	4	17	6	5
Höhe (in m)	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
Anzahl der Türen	1	1	2	2	1	2	1	2	1	1	1	1	1	2	1	2	1	3	1	1
Breite pro Tür (in m)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Angaben zum Inventar im Raum																				
Anzahl Regale	4	3	3	5	4	7	4	2	2	3	4	3	4	8	3	2	1	4	2	2
Anzahl Schränke	2	1	2	2	4	5	3	2	2	1	2	3	1	4	2	3	1	3	2	1
Anzahl Stühle	5	5	7	2	7	16	6	4	3	2	4	6	8	15	3	6	5	16	6	8
Anzahl Tische	2	2	4	2	7	12	3	3	2	2	4	3	5	14	3	4	3	11	3	5
Anzahl Computer	3	2	5	2	10	14	5	2	1	0	3	4	3	10	5	3	2	8	6	5
Anzahl Monitore	3	2	10	4	16	14	5	3	1	0	4	5	3	10	6	3	2	10	6	7

Aufgaben zum Text 1: „Renovierung eines Altenpflegeheimes“:

1. Aufgabe:

Raum 0.06 wird komplett neu renoviert. Berechnen Sie, wie viel m² rutschfester Bodenbelag verlegt werden müssen.

2. Aufgabe:

Berechnen Sie, wie viel m Bodenleisten im Raum 0.06 verlegt werden müssen. Beachten Sie dabei, dass die Breite der Türen abgezogen werden muss (vgl. Tabelle)!

3. Aufgabe:

Berechnen Sie, wie groß die Fläche ist, die von der Firma Renofix GmbH für das Streichen der Wände in Raum 0.06 in Rechnung gestellt wird. Beachten Sie dabei den Abschnitt (2) des Vertragsauszuges!

4. Aufgabe:

Raum 2.04 (im 2. Obergeschoss) wird ebenfalls komplett renoviert. Er hat eine Grundfläche von 600 m². Berechnen Sie die Kosten für die Verlegung des rutschfesten Bodenbelages.

5. Aufgabe:

In Raum 2.04 müssen insgesamt 58 m an Bodenleisten verlegt werden (nach Abzug der Türbreiten). Berechnen Sie, wie teuer die Verlegung der Bodenleisten ist.

6. Aufgabe:

In Raum 2.04 wird nach der Berechnungsvorschrift (2) aus dem Arbeitsvertrag mit einer Wandfläche von 176 m² kalkuliert. Wie teuer ist die Streichung des Raumes?

Text 2: Der ambulante Pflegedienst

Der Altenpfleger Thorsten Müller, der in Ettlingen wohnt, ist jeden Tag in der Woche mit dem Dienstauto (Opel, 60 PS, Diesel-Motor) unterwegs. In der 23. Kalenderwoche im Jahr 2015 hatte Herr Müller an einem Tag eine besonders stressige Tour:

Zu Beginn seines Dienstes um 06:30 Uhr fährt Herr Müller von seiner Dienststelle in Ettlingen nach Reichenbach, weil er dort bei einem älteren Herrn den Blutzucker messen und Insulin spritzen muss. Im Anschluss fährt er von Reichenbach nach Langensteinbach weiter, da er dort bei einem weiteren Kunden eine kleine Toilette durchzuführen hat. In Langensteinbach hat Herr Müller noch zwei weitere Kunden zu betreuen und bleibt dort bis 10:30 Uhr. Gegen 11:00 Uhr macht sich Herr Müller wieder auf den Weg zu seiner Dienststelle. Er muss auf dem Weg zurück zur Dienststelle allerdings noch einen kurzen Hausarztbesuch in Ittersbach wahrnehmen, um ein Rezept für eine Klientin abzuholen. Zwischen 12:00 Uhr und 13:00 Uhr muss Herr Müller noch nach Grünwettersbach und anschließend nach Kleinsteinbach, um dort Klienten zu betreuen.

Aufgaben zum Text 2:

7. Aufgabe:

Wie viele km hat Herr Müller zwischen 06:30 - 10:30 Uhr mit seinem Auto zurückgelegt?

8. Aufgabe:

Wie viele km hat Herr Müller zwischen 11:00 – 12:00 Uhr mit seinem Auto zurückgelegt?

9. Aufgabe:

Herr Müller fährt im Durchschnitt 100 km/h. Wie viel Liter Kraftfahrtstoff hat Herr Müller auf dem Weg zwischen Grünwettersbach und Kleinsteinbach verbraucht?

Tabelle 72: Entfernungen deutscher Städte/Dörfer in km (Landstraßen)

Alle Angaben in km	Langensteinbach	Reichenbach	Spielberg	Mutschelbach	Auerbach	Ettlingen	Ittersbach	Grünwettersbach	Kleinsteinbach
Langensteinbach		3,2	1,6	5,1	2,9	9,7	1,1	7,0	7,6
Reichenbach	3,2		4,8	9,0	5,5	7,0	4,2	6,0	8,4
Spielberg	1,6	4,8		7,0	4,8	12,2	0,8	8,9	9,5
Mutschelbach	5,1	9,0	7,0		7,5	17,5	6,2	9,1	2,7
Auerbach	2,9	5,5	4,8	7,5		12,0	4,0	9,4	9,9
Ettlingen	9,7	7,0	12,2	17,5	12,0		10,9	8,3	16,2
Ittersbach	1,1	4,2	0,8	6,2	4,0	10,9		8,1	8,6
Grünwettersbach	7,0	6,0	8,9	9,1	9,4	8,3	8,1		8,9
Kleinsteinbach	7,6	8,4	9,5	2,7	9,9	16,2	8,6	8,9	

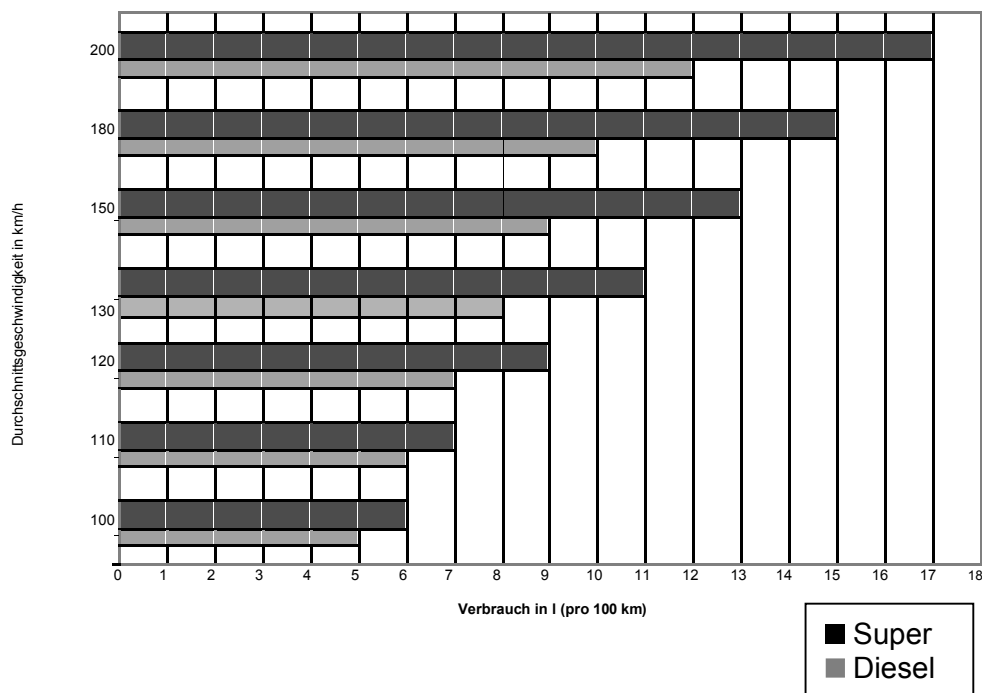


Abbildung 23: Kraftfahrstoffverbrauch in L (pro 100 km) in Abhängigkeit der Durchschnittsgeschwindigkeit

Text 3: Diabetes und Alkohol

In einer Cocktailbar werden Cocktails in den Größen klein (300 ml), mittel (400 ml) und groß (500 ml) verkauft. An einem Abend erfährt Frau Maier, welche seit einem Jahr an einem Diabetes mellitus Typ II leidet, nach einem längeren Gespräch mit dem Barkeeper die Zutaten für ihren Lieblingscocktail „Mai Tai“. Der Barkeeper zeigt ihr die Zutatenliste für die Cocktails (ein

Auszug davon ist in der Tabelle abgebildet). Frau Maier muss aufgrund ihrer Diabetes mellitus Erkrankung darauf achten, dass Sie nicht zu viel Alkohol zu sich nimmt, damit sie nicht in die Unterzuckerung gerät. Daher möchte Frau Maier nun herausfinden, wie viel Alkohol in einem „Mai Tai“ enthalten ist. Im Internet findet sie dazu das unten abgebildete Diagramm, das den Anteil an reinem Alkohol in verschiedenen alkoholischen Getränken angibt. Bearbeiten Sie die Aufgaben auf der folgenden Seite.

Tabelle 73: Auszug aus der Zutatenliste für Cocktails

Lousiana		Long Island Ice Tea		Mai Tai		Tequila Sunrise	
Orangensaft	1/6	Cola	2/5	Rum	1/6	Orangensaft	5/8
Pfirsichnektar	1/6	Gin	1/15	Gin	1/4	Tequila	1/6
Whiskey	2/5	Orangensaft	2/15	Zitronensaft	1/6	Zitronensaft	1/8
Gin	1/6	Rum	1/10	Ananassaft	1/3	Grenadine	1/12
Zitronensaft	1/10	Tequila	2/15	Pfirsichsaft	1/20		
		Wodka	1/10	Cola	1/30		
		Zitronensaft	1/15				

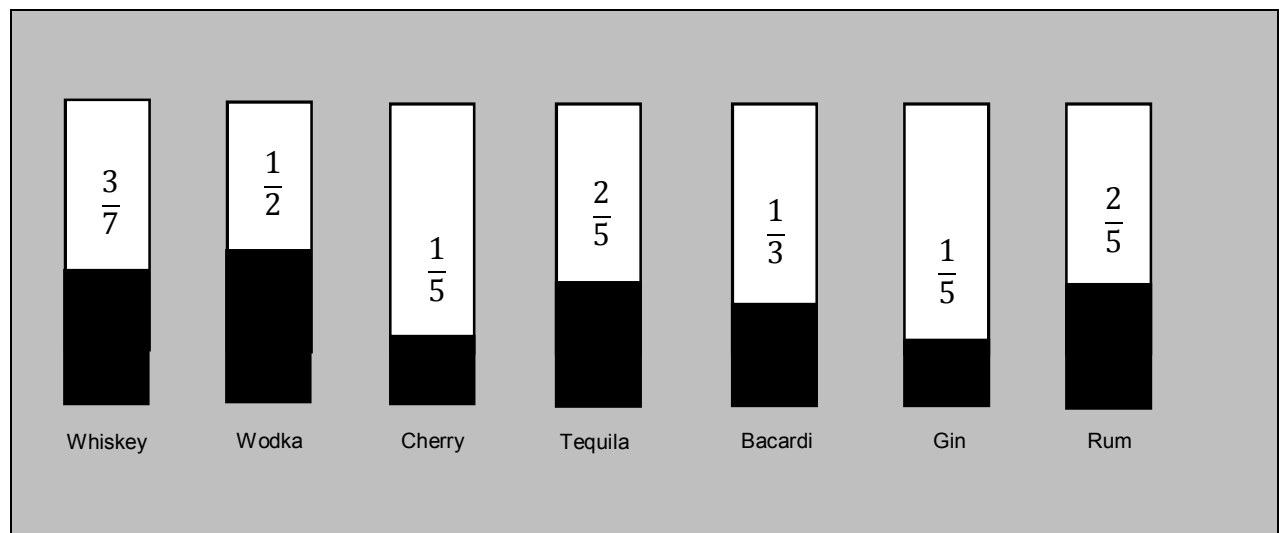


Abbildung 24: Anteil an reinem Alkohol in verschiedenen alkoholischen Getränken

10. Aufgabe:

Frau Maier hat an dem Abend in der Cocktailbar zwei kleine Mai Tai getrunken. Wieviel ml sind das?

11. Aufgabe:

Zu welchem Anteil besteht ein Mai Tai aus Fruchtsaft?

12. Aufgabe:

Geben Sie den Alkoholgehalt von Rum im Mai Tai in % an.

Text 4: Die Reinigungsfirma

Die Reinigungsfirma Cleanfix GmbH betreibt zwei weit voneinander entfernte Zweigstellen. Insgesamt arbeiten 8 Reinigungsfachkräfte in der Firma. Die Hauptaufgabe der Reinigungsfirma besteht in der hygienischen Säuberung der Böden in den Räumlichkeiten von Pflegeeinrichtungen. Auf diese Weise kann die Reinigungsfirma mit den Mitarbeitern einer Zweigstelle pro Tag 200 m² Böden reinigen. Mit dem derzeitigen Lagerbestand an speziellen Reinigungsmitteln kann die Firma noch insgesamt für 12 Arbeitstage pro Zweigstelle die Böden der Pflegeeinrichtungen reinigen. Kurzfristig erhält die Reinigungsfirma zusätzlich den Auftrag, in beiden Pflegeeinrichtungen auch die Büroböden zu reinigen. So müssen in der ersten Pflegeeinrichtung insgesamt 300 m² Böden und in der zweiten Pflegeeinrichtung 400 m² Böden pro Tag gereinigt werden.

Aufgabe zum Text 4:**13. Aufgabe:**

Nach wie vielen Arbeitstagen ist der Lagerbestand an der zweiten Zweigstelle aufgebraucht?

Aufgabentext zu Aufgabe 14:

Herr Weinfeld unterzog sich vor einer Woche einer Hüftoperation. Nach einer erfolgreichen Genesung konnte er planmäßig entlassen werden. Herr Weinfeld fährt nach seinem Krankenhausaufenthalt mit einem Taxi vom Klinikum Karlsruhe zu seiner Wohnung. Die Taxifahrt kostet 3 € zuzüglich 1,50 € pro gefahrenen km. Als Herr Weinfeld an seiner Wohnung ankommt, verlangt der Taxifahrer einen Preis von 15,75 €.

14. Aufgabe:

Wie viele km wohnt Herr Weinfeld vom Klinikum entfernt?

Aufgabentext zu Aufgabe 15:

In einer Firma für Medizinprodukte werden verschiedene Produkte hergestellt. Unter anderem Kompressen, Mullbinden und Wundpflaster. Damit für die Kunden immer genügend Materialien zur Verfügung stehen, hat sich die Firma entschieden, ein kleines Lager zu halten, um die Überproduktion kurzfristig lagern zu können. Am Anfang der 34. Kalenderwoche (KW) (also Montagmorgen) sind in diesem Lager 323 Packungen mit Wundpflaster, 409 Packungen mit Mullbinden sowie 89 Packungen mit Kompressen. Gegen Ende der 34. KW (also Freitagabend) werden die Produkte vom Transportunternehmen „Maier-Transporte“ abgeholt und zu den Kunden geschickt – je nach Auftragslage werden etwa 100 Packungen mit Kompressen, 350 Packungen mit Wundpflaster und 500 Packungen mit Mullbinden pro Woche abgeholt. Bevor die Firma Maier Transporte am Freitagnachmittag die Produkte abholt, hat das Lager der Fabrik einen Bestand von 634 Packungen mit Mullbinden, 169 Packungen mit Kompressen und 526 Packungen mit Wundpflaster.

15. Aufgabe:

Wie viele Packungen mit Mullbinden wurden in der 34. KW hergestellt?

Alle Texte und Aufgaben des CareMaTex wurden nur inhaltlich an den Kontext des Altenpflegeberufs angepasst. Die mathematischen Operatoren und Operationen blieben – bis auf Text 2: Der ambulante Pflegedienst – von der Überarbeitung unberührt. Hier mussten die Angaben zu den Wegstrecken numerisch geändert werden. Daraus ergab sich bei den zugehörigen Aufgaben eine Änderung der Lösungswege, welche in Tabelle 74 dargestellt werden sollen.

Tabelle 74: Lösungswege zu den Aufgaben 6 - 8

Aufgabe	IALS-Stufe	Mathematische Anforderungen	Rechenoperationen
6	Stufe 3	Arithmetik: Addition	$7,0 + 3,2$
7	Stufe 4	Arithmetik: Addition	$1,1 + 10,9$
8	Stufe 5	• Schritt: Tabelle lesen	8,9 km
		• Schritt: Diagramm lesen	$5 \frac{L}{100 km}$

		<ul style="list-style-type: none"> Schritt: Arithmetik: Multiplikation und Division 	$5L \frac{8,9 km}{100 km}$
--	--	--	----------------------------

11.2 Gütekriterien des CareMaTex

Im Folgenden soll auf die Gütekriterien Objektivität, Reliabilität und Validität des CareMaTex eingegangen werden.

11.2.1 Kriterium der Objektivität

Durchführungsobjektivität: In Bezug auf den CareMaTex wird die Durchführungsobjektivität auf verschiedene Weisen erreicht: Für die neue Testversion wurde das Testmanual des TeMaTex überarbeitet und angepasst. Ein entsprechendes Instruktionsblatt zur Durchführung des Tests liegt vor, welches den Probanden vor der Durchführung des Tests vom Testleiter vorgelesen wird. Somit werden die Probanden über den Ablauf informiert. Die Testbögen sind standardisiert, sodass für jede Testperson dieselben Testbedingungen bestehen. Die Testzeit beträgt inklusive Testinstruktion durch den Testleiter 45 Minuten. Es sind keine Hilfsmittel wie Taschenrechner erlaubt. Das Testmaterial besteht aus einem Testbogen. Die Probanden tragen die Lösungen zu den Testaufgaben in den Testbogen ein. Hierfür liegen extra farbig hinterlegte Felder auf dem Testbogen vor, in welche die Zahl oder der Betrag eingetragen werden müssen.

Auswertungsobjektivität: Die Auswertungsobjektivität wird im Rahmen des CareMaTex in Anlehnung an den TeMaTex durch folgende Regeln erreicht: Löst ein Proband ein Item richtig, wird dies mit einem Punkt bewertet. Es gelten nur diejenigen Antworten als richtig, welche in das vorgesehene Lösungsfeld auf dem Testbogen eingetragen werden. Die in die Lösungsfelder eingetragenen Lösungen werden im Anschluss in eine eigens dafür erstellte Excel-gestützte Eingabemaske eingetragen, in der die Rohdaten automatisch zu richtig oder falsch beantworteten Aufgaben umcodiert werden.

11.2.2 Kriterium der Reliabilität

Innere Konsistenz: Insgesamt wurden $n = 253$ Probanden in die Analyse einbezogen. Zur Bestimmung der inneren Konsistenz wurde der Andrichs-Reliabilitäts-Koeffizient verwendet. Die innere Konsistenz des vorliegenden Tests liegt bei $r = 0,70419$.

11.2.3 Kriterium der Validität

Inhaltsvalidität: Die Aufgaben des CareMaTex wurden ausschließlich auf der semantischen Ebene an den Kontext der Altenpflege angepasst, sodass sichergestellt werden konnte, dass die relevanten mathematischen Grundlagenbereiche wie Arithmetik, Prozentrechnung und Bruchrechnung mithilfe der 15 Items abgedeckt werden.

Konstruktvalidität: Für den CareMaTex wurde mithilfe der Software-Programme WINMIRA und RUMM 2030 eine Modellgeltung nachgewiesen. Zur ausführlichen Beschreibung des Verfahrens zur Ermittlung der Konstruktvalidität des CareMaTex sei auf Kapitel 10 verwiesen.

11.3 Zusammenfassende Betrachtung des elften Kapitels

In Kapitel 11 wurde der CareMaTex ausführlich beschrieben. Insgesamt umfasst der neue Test 15 Items, welche an den Kontext der Altenpflege angepasst wurden. Es konnte nachgewiesen werden, dass der CareMaTex eine ausreichende Durchführungs- und Auswertungsobjektivität, interne Konsistenz und Inhalts- und Konstruktvalidität aufweist.

V. Diskussion

Im Folgenden werden die in Kapitel 8.2.1 und 10.2.1 ermittelten Ergebnisse diskutiert und interpretiert. Im Zuge dessen werden die zuvor aufgestellten Hypothesen überprüft und es wird versucht, die Forschungsfragen zu beantworten. Der Diskussionsteil wird in Kapitel 13 mit einem Fazit abgeschlossen.

12 Diskussion der Ergebnisse

In Kapitel 12.1 werden die Ergebnisse aus Kapitel 8.2.1 und in Kapitel 12.2 die Ergebnisse aus Kapitel 10.2.1 interpretiert und daraus Schlussfolgerungen gezogen. In Kapitel 12.3 wird auf die Forschungsfragen und Hypothesen dieser Arbeit eingegangen.

12.1 Diskussion der Ergebnisse der statistischen Analysen in Bezug auf den TeMaTex

Die Rasch-Analyse mit dem Statistikprogramm WINMIRA ergab, dass keine Modellverletzung vorliegt. Vor diesem Hintergrund ist es nun möglich, erste Berechnungen hinsichtlich des mathematischen Textverständnisses der Altenpflegeschüler durchzuführen.

Diskussion der Ergebnisse aus der Analyse mit WINMIRA: Im Folgenden werden die Ergebnisse aus den Rasch-Analysen mit der Software WINMIRA mit den Teilergebnissen der Untersuchung von Jordan 2011 verglichen. Aus der Studie von Jordan werden als Referenzwerte die Prozentrangnormen für die unterschiedlichen Schulabschlüsse angegeben, die in Tabelle 75 zusammengestellt sind.

Tabelle 75: Prozentrangnormen Schulabschlüsse (nach Jordan 2011a, S. 357)

	Hauptschulabschluss	Realschulabschluss	Fachhochschulabschluss/Abitur
0	4 %	2 %	1 %
1	9 %	3 %	2 %
2	14 %	4 %	3 %
3	18 %	6 %	4 %
4	26 %	7 %	6 %
5	32 %	11 %	6 %

6	40 %	13 %	7 %
7	47 %	19 %	10 %
8	56 %	26 %	12 %
9	62 %	34 %	16 %
10	72 %	43 %	23 %
11	78 %	54 %	30 %
12	84 %	66 %	39 %
13	89 %	75 %	48 %
14	95 %	84 %	59 %
15	96 %	90 %	70 %
16	98 %	94 %	81 %
17	100 %	99 %	90 %
18	100 %	100 %	100 %

In der linken Spalte der Tabelle stehen die theoretisch erreichbaren Gesamtpunktzahlen (0 bis 18 Punkte). In der zweiten, dritten und vierten Spalte sind die Prozentrangnormen zu den jeweiligen Schulabschlüssen angegeben. Die grau hinterlegten Zellen stellen den Bereich um den Mittelwert \pm Standardabweichung dar. Die Prozentrangnormen können als Grundlage für die Leistungsbeurteilung in Form einer Notengebung gesehen werden (vgl. Ingenkamp & Lissmann 1997, S. 64). Tabelle 76 verdeutlicht die Leistungsbeurteilung anhand der eingeteilten Prozenträge.

Tabelle 76: Übersicht Leistungsbeurteilung (nach Schneider et al. 2007, S. 20)

Prozentrang	entspricht einer	Note
96 - 100	ausgezeichneten Leistung	sehr gut
76 - 95	überdurchschnittlichen Leistung	gut
51 - 75	durchschnittlichen Leistung (oberer Durchschnitt)	befriedigend
26 - 50	durchschnittlichen Leistung (unterer Durchschnitt)	ausreichend
11 - 25	unterdurchschnittlichen Leistung	mangelhaft
6 - 10	schwachen Leistung	ungenügend
0 - 5	sehr schwachen Leistung	

Tabelle 77 können die nach Schulabschlüssen sortierten Leistungen der Probanden entnommen werden. Die Tabelle liest sich wie folgt: Ein Proband mit Hauptschulabschluss erreicht eine Gesamtpunktzahl von 4 Punkten. Nach den Tabellen 76 und 77 wird dieser in den Prozentrang zwischen 26

und 50 % eingestuft und erhält somit die Note „ausreichend“. Die Gesamtpunktzahl von 4 Punkten erreichten von den 125 Probanden 16. Insgesamt schneiden die Probanden mit Realschulabschluss und Fachhochschulreife/Abitur am schlechtesten ab. 40 % der Probanden mit Realschulabschluss erhalten die Note „ungenügend“ und 29,1 % die Note „mangelhaft“. Bei den Probanden mit Fachhochschulreife/Abitur erhalten 37,3 % die Note „ungenügend“ und 29,8 % die Note „mangelhaft“. Das schlechte Abschneiden der Gruppe „Fachhochschulreife/Abitur“ könnte daran liegen, dass diese Gruppe in der Gesamtstichprobe der Untersuchung unterrepräsentiert ist und es aufgrund dessen zu Verzerrungen kommt. Bei den Probanden mit Hauptschulabschluss sieht es wie folgt aus: 29,6 % der Probanden mit Hauptschulabschluss erhalten die Note „ungenügend“ und 14,4 % die Note „mangelhaft“.

Tabelle 77: WINMIRA: Schülerleistungen in Bezug auf den TeMaTex

Gesamtpunktwerte	Hauptschulabschluss			Realschulabschluss			Fachhochschulabschluss/Abitur		
0	4 %	20	ungenügend	2 %	11	ungenügend	1 %	1	ungenügend
1	9 %	17		3 %	9		2 %	1	
2	14 %	14		4 %	6		3 %	3	
3	18 %	14		6 %	24		4 %	3	
4	26 %	16	ausreichend	7 %	24	mangelhaft	6 %	3	ungenügend
5	32 %	14		11 %	14		6 %	4	
6	40 %	8		13 %	19		7 %	8	
7	47 %	10		19 %	20		10 %	2	
8	56 %	5	befriedigend	26 %	20	ausreichend	12 %	7	mangelhaft
9	62 %	2		34 %	11		16 %	11	
10	72 %	0		43 %	6		23 %	2	
11	78 %	1	gut	54 %	10	befriedigend	30 %	4	ausreichend
12	84 %	0		66 %	1		39 %	4	
13	89 %	1		75 %	3		48 %	4	
14	95 %	2	sehr gut	84 %	1	gut	59 %	3	befriedigend
15	96 %	0		90 %	0		70 %	2	
16	98 %	1		94 %	3	sehr gut	81 %	4	gut
17	100 %	0		99 %	0		90 %	1	
18	100 %	0		100 %	0		100 %	0	

Tabelle 78 zeigt eine Erweiterung der in Kapitel 7.1 dargestellten Kompetenzstufeneinteilung. Hinzugefügt wurde die Kompetenzstufe 0. Diese Erweiterung resultierte aus den Berechnungen der Kompetenzstufeneinteilung mit einer Lösungswahrscheinlichkeit von 70 %. Auch Jordan (2011) verwendete in seinen Analysen diesen Wert. Dies ermöglicht anhand der Ergebnisse aus den Analysen mit WINMIRA, die Ergebnisse aus dieser Untersuchung mit den Ergebnissen aus der Studie von Jordan zu vergleichen.

Tabelle 78: Beschreibung der Kompetenzstufen (modifiziert nach Jordan 2011a, S. 351)

Kompetenzstufe 0	Probanden, welche dieser Stufe zugeordnet werden, können nicht mit einer Wahrscheinlichkeit von mindestens 70 % die leichteste Aufgabe der Kompetenzstufe I lösen, das heißt, dass die Probanden vermutlich nicht in der Lage sind, aus dem Textmaterial zwei natürliche Zahlen zu identifizieren, um anschließend eine einfache Addition oder Multiplikation durchführen zu können.
Kompetenzstufe I	Bei Aufgaben der ersten Kompetenzstufe wird von den Probanden verlangt, dass sie aus dem Textmaterial zwei natürliche Zahlen identifizieren. Anschließend muss eine Addition oder Multiplikation durchgeführt werden. Die Zahlen sind im Material leicht zu finden und die Rechenoperationen sind ebenfalls einfach. Im mathematischen Lösungsprozess muss nur eine einzige Rechenoperation durchgeführt werden.
Kompetenzstufe II	Bei Aufgaben der zweiten Kompetenzstufe müssen aus dem Textmaterial zwei oder mehr Zahlen identifiziert werden. Aufgrund vieler konkurrierender Informationen sind die benötigten Zahlen jedoch nicht direkt ersichtlich, sodass zum Auffinden der benötigten Informationen ein umfassendes Textverständnis notwendig ist. Die mathematischen Operationen sind zudem schwerer als bei Aufgaben der ersten Kompetenzstufe. So werden beispielsweise Kenntnisse der Bruchrechnung, Prozentrechnung und Dezimalbruchrechnung verlangt, während bei Aufgaben der ersten Kompetenzstufe ausschließlich mit natürlichen Zahlen gerechnet wird. Meist muss im Lösungsprozess ein einziger Rechenschritt durchgeführt werden. Es gibt jedoch auch Aufgaben, bei denen mehr als ein Rechenschritt durchzuführen ist.
Kompetenzstufe III	Aufgaben der dritten Stufe zeichnen sich dadurch aus, dass sämtliche Textmaterialien (Fließtext, Diagramm, Tabelle) zur Bearbeitung der Aufgaben hinzugezogen werden müssen. Mehrere Informationen müssen miteinander verknüpft werden. Die Suche nach den benötigten Informationen wird dadurch erschwert, dass in den Texten eine oder mehrere konkurrierende Informationen gegeben sind. Für den Lösungsprozess sind in der Regel mehr als zwei Rechenschritte notwendig.

Im Rahmen der Rasch-Analyse lassen sich die theoretisch erreichbaren Gesamtscores den in Tabelle 78 beschriebenen Kompetenzstufen zuordnen und darstellen. In der linken Spalte von Tabelle 79 sind die Kompetenzstufen ersichtlich. Die zweite Spalte gibt Auskunft über die Punktzahlen,

welche im TeMaTex erreicht werden können. Die dritte und vierte Spalte geben Informationen über die Personenparameter und die dazugehörigen Itemparameter.

Tabelle 79: WINMIRA: Kompetenzstufen in Bezug auf den TeMaTex

Kompetenzstufe	Punkte	Personenparameter	Itemparameter
0	0	-4,498	-5,345
	1	-3,261	-4,108
	2	-2,604	-3,451
	3	-2,113	-2,960
Schwelle der Kompetenzstufe 0 zu I			
I	4	-1,699	-2,546
	5	-1,328	-2,175
	6	-0,984	-1,831
	7	-0,657	-1,504
	8	-0,340	-1,187
	9	-0,029	-0,876
	10	0,283	-0,564
Schwelle der Kompetenzstufe I zu II			
II	11	0,601	-0,246
	12	0,932	0,085
	13	1,287	0,440
Schwelle der Kompetenzstufe II zu III			
III	14	1,678	0,831
	15	2,124	1,277
	16	2,658	1,811
	17	3,365	2,518
	18	4,652	3,805

Die Tabelle lässt sich wie folgt interpretieren: Ein Proband erreicht in dem Test die Gesamtpunktzahl von 10 Punkten, sodass er eine Personenfähigkeit von 0,283 hat. Durch die Formel zur Bestimmung der Differenz in Bezug auf die Personen- und Itemparameter (siehe Gl. 8) ergibt sich ein Itemparameter von -0,564. Dies besagt, dass der Proband nur Items mit einer Itemschwierigkeit von maximal -0,564 erfolgreich bearbeiten kann. Da aber das leichteste Item der Kompetenzstufe II eine Itemschwierigkeit von -0,246 aufweist, wird er dieses Item nicht mit einer 70-prozentigen Wahrscheinlichkeit lösen können. In Kompetenzstufe 0 werden alle Probanden mit einer Gesamtpunktzahl von 0 Punkten bis einschließlich 3 Punkten eingeordnet. In die erste Kompetenzstufe werden alle Probanden mit einem Gesamtpunktescore von 4 Punkten bis einschließlich 10 Punkten eingeteilt. Bei Kompetenzstufe II sind es die Probanden mit einem Gesamtpunktescore

von 11 Punkten bis einschließlich 13 Punkten und bei Kompetenzstufe III sind es Probanden mit einem Gesamtpunktescore von 14 Punkten bis 18 Punkten.

Basierend auf der beschriebenen Kompetenzstufeneinteilung und der Häufigkeitsverteilung der Gesamtscores können nun Aussagen hinsichtlich der Kompetenzniveaus der Altenpflegeschüler – sortiert nach den Schulabschlüssen – getroffen werden. Die Ergebnisse zeigen, dass circa 52 % der Hauptschüler die Kompetenzstufe 0; 44 % der Hauptschüler die Stufe I; 1,6 % der Hauptschüler die Kompetenzstufe II und 2,4 % der Hauptschüler die Kompetenzstufe III erreicht haben. Jordan (2011) erhielt ein ähnlich alarmierendes Ergebnis, wonach circa 25 % der befragten Hauptschüler (n = 467) lediglich die Kompetenzstufe 0 und 60 % die Stufe I erreichten. Nur 11,8 % erreichten die Kompetenzstufe II und nur 3,2 % die dritte Stufe (vgl. Jordan 2011a, S. 353). Die Realschüler schneiden in der vorliegenden Untersuchung vergleichsweise besser ab. Etwa 27,5 % der Realschüler erreichen die Kompetenzstufe 0 und 62,6 % die Stufe I, 7,7 % die Stufe II und 2,2 % die Stufe III. Bei Jordan (2011) erreichten circa 6,6 % der Realschüler (n = 640) die Kompetenzstufen 0 und etwa 55,4 % die Stufe I, circa 25 % erreichten die Stufe II und 13 % der Realschüler die Kompetenzstufe III (vgl. Jordan 2011a, S. 353 f.). Im Vergleich zu den Hauptschülern und Realschülern schneiden in der vorliegenden Untersuchung die Schüler mit Fachhochschulreife beziehungsweise Fachabitur oder Abitur besser ab. Circa 12 % der Schüler erreichen die Kompetenzstufe 0 und 55,2 % die Stufe I; 17,9 % die Kompetenzstufe II und 14,9 % die Stufe III. In der Untersuchung von Jordan (2011) schneidet diese Schülergruppe (n = 566) deutlich besser ab. Ungefähr 25 % der Schüler erreichen jeweils die Kompetenzstufe I und II und circa 50 % die Stufe III (vgl. Jordan 2011a, S. 354).

Aus den Ergebnissen der Analyse mit der Software WINMIRA wird ersichtlich, dass sich in Bezug auf das Kompetenzniveau große Unterschiede zwischen den Schulabschlüssen bemerkbar machen, was in Tabelle 80 dargestellt wird. Die Ergebnisse lassen vermuten, dass mit Erwerb eines höheren

Schulabschlusses auch eine Steigerung des mathematischen Textverständnisses einhergeht.

Tabelle 80: Übersicht Kompetenzniveau sortiert nach Schulabschlüssen – Teil 1

Kompetenzstufen	Hauptschulabschluss	Realschulabschluss	Fachhochschulreife/ Abitur
Kompetenzstufe 0	52 %	27,5 %	12 %
Kompetenzstufe I	44 %	62,6 %	55,2 %
Kompetenzstufe II	1,6 %	7,7 %	17,9 %
Kompetenzstufe III	2,4 %	2,2 %	14,9 %

Diskussion der Ergebnisse aus der Analyse mit RUMM 2030: Die Richtlinien zur Auswertung des TeMaTex geben vor, dass für jede richtige Bearbeitung einer Aufgabenstellung des TeMaTex ein Punkt vergeben wird. Somit können die Probanden nach der Elimination von Item 2, 5, 13 und 15 einen Gesamtpunktescore von 14 Punkten erreichen. Aufgrund der vorliegenden Modellgeltung ist es erlaubt, die einzelnen Punkte zu einem Gesamtpunktescore zu summieren. Im Rahmen der Rasch-Analyse wurden zudem die Itemparameter der verbliebenen Items und die Personenparameter für die Summenscores im Bereich von 0 bis 14 Punkten geschätzt.

Im weiteren Verlauf werden die theoretisch erreichbaren Gesamtscores aufgrund der Rasch-Analyse mit der Software RUMM 2030 in eine der vier Kompetenzstufen eingeordnet. Aus der Tabelle 81 kann diese Zuordnung entnommen werden.

Tabelle 81: RUMM 2030: Kompetenzstufen in Bezug auf den TeMaTex

Kompetenzstufe	Punkte	Personenparameter	Itemparameter
0	0	-3,957	-4,804
	1	-3,056	-3,903
	2	-2,347	-3,194
	3	-1,790	-2,637
Schwelle der Kompetenzstufe 0 zu I			
I	4	-1,296	-2,143
	5	-0,836	-1,683
	6	-0,401	-1,248
	7	0,015	-0,832
Schwelle der Kompetenzstufe I zu II			
II	8	0,420	-0,427
	9	0,828	-0,019

	10	1,259	0,412
Schwelle der Kompetenzstufe II zu III			
III	11	1,740	0,893
	12	2,316	1,469
	13	3,080	2,233
	14	4,097	3,250

In die Kompetenzstufe 0 werden alle Probanden mit einem Gesamtpunktscore von 0 bis einschließlich 3 Punkten eingeordnet. In die erste Kompetenzstufe werden alle Probanden mit einem Gesamtpunktscore von 4 Punkten bis einschließlich 7 Punkten eingeteilt. Bei Kompetenzstufe II sind es die Probanden mit einem Gesamtpunktscore von 8 Punkten bis einschließlich 10 Punkten und bei Kompetenzstufe III sind es Probanden mit einem Gesamtpunktscore von 11 Punkten bis 14 Punkten. Tabelle 82 zeigt die Verteilung der Schülerleistungen.

Tabelle 82: RUMM 2030: Schülerleistungen in Bezug auf den TeMaTex

Ge- samt- punkt- werte	Hauptschulab- schluss	Realschulab- schluss	Fachhochschulab- schluss/Abitur
0	20	11	2
1	23	16	0
2	17	14	4
3	20	24	8
4	14	29	3
5	16	26	8
6	6	24	6
7	4	13	11
8	0	8	5
9	2	10	4
10	0	2	3
11	2	2	4
12	0	2	5
13	0	1	4
14	1	0	0
	n = 125	n = 182	n = 67

Die Ergebnisse der Analyse zeigen, dass circa 64 % der Hauptschüler die Kompetenzstufe 0, 32 % der Hauptschüler die Stufe I, 1,6 % der Hauptschüler die Kompetenzstufe II und 2,4 % der Hauptschüler die Kompetenzstufe III erreicht haben. Die Realschüler schneiden wieder vergleichsweise

besser ab. Etwa 35,7 % der Realschüler erreichen die Kompetenzstufe 0 und 50,5 % die Stufe I, etwa 11 % die Stufe II und 2,8 % die Stufe III. Im Vergleich zu den Hauptschülern und Realschülern schneiden die Schüler mit Fachhochschulreife beziehungsweise Fachabitur oder Abitur auch wieder besser ab. Circa 20,9 % der Schüler erreichen die Kompetenzstufe 0 und 41,8 % die Stufe I; 17,9 % die Kompetenzstufe II und 19,4 % die Stufe III. Aus den Ergebnissen der Analyse mit der Software RUMM 2030 wird ersichtlich, dass sich auch hier große Unterschiede in Bezug auf das Kompetenzniveau zwischen den Schulabschlüssen ergeben, was in Tabelle 83 dargestellt wird.

Tabelle 83: Übersicht Kompetenzniveau sortiert nach Schulabschlüssen - Teil 2

Kompetenzstufen	Hauptschulabschluss	Realschulabschluss	Fachhochschulreife/ Abitur
Kompetenzstufe 0	64 %	35,7 %	20,9 %
Kompetenzstufe I	32 %	50,5 %	41,8 %
Kompetenzstufe II	1,6 %	11 %	17,9 %
Kompetenzstufe III	2,4 %	2,8 %	19,4 %

Diskussion der Ergebnisse aus der Analyse mit eRm: Im Folgenden werden die theoretisch erreichbaren Gesamtscores, aufgrund der Rasch-Analyse mit eRm, in eine der vier Kompetenzstufen eingeordnet. Tabelle 84 kann diese Zuordnung entnommen werden.

Tabelle 84: eRm: Kompetenzstufen in Bezug auf den TeMaTex

Kompetenzstufe	Punkte	Personenparameter	Itemparameter
0	0	-4,514	-5,361
	1	-3,607	-4,454
	2	-2,768	-3,615
	3	-2,210	-3,057
	4	-1,761	-2,608
Schwelle der Kompetenzstufe 0 zu I			
I	5	-1,369	-2,216
	6	-1,011	-1,858
	7	-0,673	-1,520
	8	-0,348	-1,195
	9	-0,027	-0,874
	10	0,293	-0,554
Schwelle der Kompetenzstufe I zu II			
II	11	0,622	-0,225
	12	0,966	0,119

	13	1,336	0,489
Schwelle der Kompetenzstufe II zu III			
III	14	1,747	0,900
	15	2,224	1,377
	16	2,819	1,972
	17	3,701	2,854
	18	4,652	3,805

In die Kompetenzstufe 0 werden alle Probanden mit einem Gesamtpunktscore von 0 bis einschließlich 4 Punkten eingeordnet. In die erste Kompetenzstufe werden alle Probanden mit einem Gesamtpunktscore von 5 Punkten bis einschließlich 10 Punkten eingeteilt. Bei Kompetenzstufe II sind es die Probanden mit einem Gesamtpunktscore von 11 Punkten bis einschließlich 13 Punkten und bei Kompetenzstufe III sind es Probanden mit einem Gesamtpunktscore von 14 Punkten bis 18 Punkten.

Die Ergebnisse zeigen, dass circa 64,8 % der Hauptschüler die Kompetenzstufe 0; 31,2 % der Hauptschüler die Stufe I, 1,6 % der Hauptschüler die Kompetenzstufe II und 2,4 % der Hauptschüler die Kompetenzstufe III erreicht haben. Die Realschüler schneiden erneut vergleichsweise besser ab. Etwa 40,6 % der Realschüler erreichen die Kompetenzstufe 0 und 49,5 % die Stufe I, 7,7 % die Stufe II und 2,2 % die Stufe III. Im Vergleich zu den Hauptschülern und Realschülern schneiden die Schüler mit Fachhochschulreife beziehungsweise Fachabitur oder Abitur auch erneut besser ab. Circa 16,5 % der Schüler erreichen die Kompetenzstufe 0 und 50,7 % die Stufe I, 17,9 % die Kompetenzstufe II und 14,9 % die Stufe III. Aus den Ergebnissen der Analyse mit eRm wird deutlich, dass sich wiederholt große Unterschiede in Bezug auf das Kompetenzniveau zwischen den Schulabschlüssen ergeben, welche Tabelle 85 dargestellt.

Tabelle 85: Übersicht Kompetenzniveau sortiert nach Schulabschlüssen - Teil 3

Kompetenzstufen	Hauptschulabschluss	Realschulabschluss	Fachhochschulreife/ Abitur
Kompetenzstufe 0	64,8 %	40,6 %	16,5 %
Kompetenzstufe I	31,2 %	49,5 %	50,7 %
Kompetenzstufe II	1,6 %	7,7 %	17,9 %
Kompetenzstufe III	2,4 %	2,2 %	14,9 %

Zusammenfassung: Die Ergebnisse aus den Berechnungen sind mit Vorsicht zu interpretieren. Das schlechte Abschneiden der Altenpflegeschüler in dieser Zielgruppe könnte auf die Testaufgaben zurückzuführen sein. Es wäre denkbar, dass die vorliegenden Testaufgaben des TeMaTex inhaltlich nicht kompatibel sind mit dem Beruf der Altenpflege. Konkret bedeutet dies, dass sich die Schüler mit einigen Testaufgaben nicht identifizieren konnten und es daher zu einem Nicht-Lösen der entsprechenden Testaufgaben kam.

Zudem ist die Kompetenzeinschätzung anhand der berechneten Kompetenzstufen begrenzt aussagekräftig, da für eine hinreichende Einschätzung der verwendete Itempool zu gering war, um alle Kompetenzstufen ausreichend mit Items abdecken zu können und valide Aussagen im Hinblick auf das Erreichen einer Kompetenzstufe machen zu können.

12.2 Diskussion der Ergebnisse der statistischen Analysen in Bezug auf den CareMaTex

Die Rasch-Analyse mit dem Statistikprogramm WINMIRA ergab, dass keine Modellverletzung vorliegt, unter Beibehaltung der gesamten Itemanzahl des Tests. Vor diesem Hintergrund ist es nun möglich, erste Berechnungen hinsichtlich des mathematischen Textverständnisses der Altenpflegeschüler durchzuführen.

Diskussion der Ergebnisse aus der Analyse mit WINMIRA: Im Rahmen der Rasch-Analyse lassen sich die theoretisch erreichbaren Gesamtscores den beschriebenen Kompetenzstufen zuordnen und darstellen. In der linken Spalte von Tabelle 86 sind die Kompetenzstufen ersichtlich. Die zweite Spalte gibt Auskunft über die Punktzahlen, welche im TeMaTex erreicht werden können. Die dritte und vierte Spalte geben Informationen über die Personenparameter und den dazugehörigen Itemparameter.

Tabelle 86: WINMIRA: Kompetenzstufen in Bezug auf den CareMaTex

Kompetenzstufe	Punkte	Personenparameter	Itemparameter
0	0	-4.554	-5,401
	1	-3.292	-4,139
	2	-2.605	-3,452
	3	-2.080	-2,927
Schwelle der Kompetenzstufe 0 zu I			
I	4	-1.630	-2,477
	5	-1.224	-2,071
	6	-0.845	-1,692
	7	-0.484	-1,331
	8	-0.134	-0,981
	9	0.209	-0,638
Schwelle der Kompetenzstufe I zu II			
II	10	0.551	-0,296
	11	0.897	0,050
	12	1.253	0,406
Schwelle der Kompetenzstufe II zu III			
III	13	1.632	0,785
	14	2.051	1,204
	15	2.546	1,699
	16	3.209	2,362
	17	4.454	3,607

In Kompetenzstufe 0 werden alle Probanden mit einem Gesamtpunktscore von 0 Punkten bis einschließlich 3 Punkten eingeordnet. In die erste Kompetenzstufe werden alle Probanden mit einem Gesamtpunktscore von 4 Punkten bis einschließlich 9 Punkten eingeteilt. Bei Kompetenzstufe II sind es die Probanden mit einem Gesamtpunktscore von 10 Punkten bis einschließlich 12 Punkten und bei Kompetenzstufe III sind es Probanden mit einem Gesamtpunktscore von 13 Punkten bis 17 Punkten.

Basierend auf der Kompetenzstufeneinteilung und der Häufigkeitsverteilung der Gesamtscores, welche in Tabelle 87 dargestellt wird, können nun – sortiert nach den Schulabschlüssen – Aussagen hinsichtlich der Kompetenzniveaus der Altenpflegeschüler getroffen werden.

Tabelle 87: WINMIRA: Schülerfähigkeit in Bezug auf den CareMaTex nach der Itemelimination

	Hauptschulabschluss	Realschulabschluss	Fachhochschulabschluss/Abitur
0	13	13	1
1	13	12	3
2	16	10	4
3	14	13	0
4	8	10	4
5	9	8	5
6	7	6	2
7	7	15	2
8	2	8	1
9	3	5	3
10	2	4	3
11	0	0	2
12	0	1	2
13	0	0	3
14	0	0	1
15	0	1	1
16	0	1	1
17	0	0	0
	n = 94	n = 107	n = 38

Es zeigt sich, dass 59,6 % der Hauptschüler die Kompetenzstufe 0, circa 38,3 % der Hauptschüler die Stufe I, 2,1 % der Hauptschüler die Kompetenzstufe II und 0,0 % der Hauptschüler die Kompetenzstufe III erreicht haben. Die Realschüler schneiden in dieser Untersuchung vergleichsweise besser ab. Etwa 44,9 % der Realschüler erreichen die Kompetenzstufe 0 und 48,6 % die Stufe I, 4,7 % die Stufe II und 1,8 % die Stufe III. Im Vergleich zu den Hauptschülern und Realschülern schneiden die Schüler mit Fachhochschulreife beziehungsweise Fachabitur oder Abitur besser ab. 21,1 % der Schüler erreichen die Kompetenzstufe 0 und 44,7 % die Stufe I, 18,4 % die Kompetenzstufe II und 15,8 % die Stufe III.

Aus den Ergebnissen der Analyse mit der Software WINMIRA wird ersichtlich, dass sich in Bezug auf das Kompetenzniveau große Unterschiede zwischen den Schulabschlüssen bemerkbar machen, was in Tabelle 88 dargestellt wird.

Tabelle 88: Übersicht über das nach Schulabschlüssen sortierte Kompetenzniveau – Teil 4

Kompetenzstufen	Hauptschulabschluss	Realschulabschluss	Fachhochschulreife/ Abitur
Kompetenzstufe 0	59,6 %	44,9 %	21,1 %
Kompetenzstufe I	38,3 %	48,6 %	44,7 %
Kompetenzstufe II	2,1 %	4,7 %	18,4 %
Kompetenzstufe III	0 %	1,8 %	15,8 %

Nach der Elimination von Item 14 und 16 (vgl. Kapitel 10.2.1.2) müssen die Kompetenzstufen neu eingeteilt werden, was Tabelle 89 zeigt.

Tabelle 89: WINMIRA: Kompetenzstufen in Bezug auf den CareMaTex nach Itemelimination

Kompetenzstufe	Punkte	Personenparameter	Itemparameter
0	0	-4.400	-5.247
	1	-3.124	-3.971
	2	-2.416	-3.263
	3	-1.865	-2.712
Schwelle der Kompetenzstufe 0 zu I			
I	4	-1.385	-2.232
	5	-0.945	-1.792
	6	-0.533	-1.380
	7	-0.139	-0.986
	8	0.242	-0.605
Schwelle der Kompetenzstufe I zu II			
II	9	0.620	-0.227
	10	1.001	0.154
Schwelle der Kompetenzstufe II zu III			
III	11	1.400	0.553
	12	1.835	0.988
	13	2.346	1.499
	14	3.024	2.177
	15	4.286	3.439

Basierend auf der Kompetenzstufeneinteilung und der Häufigkeitsverteilung der Gesamtscores können nun wiederum Aussagen hinsichtlich der Kompetenzniveaus der Altenpflegeschüler – sortiert nach den Schulabschlüssen – getroffen werden. Aufgrund der Tatsache, dass die Ergebnisse der Ausführungen nicht wesentlich von den Ergebnissen der Schülerleistungen abweichen, welche mit RUMM 2030 erzeugt wurden, wird auf eine detaillierte Darlegung der prozentualen Schülerleistungsverteilung verzichtet und auf die im folgenden Abschnitt vorgestellten Ergebnisse verwiesen.

Diskussion der Ergebnisse aus der Analyse mit RUMM 2030: Nach der Elimination von Item 14 und 16 werden im weiteren Verlauf die theoretisch erreichbaren Gesamtscores, aufgrund der Rasch-Analyse mit der Software RUMM 2030, in eine der vier Kompetenzstufen eingeordnet. Tabelle 90 kann diese Zuordnung entnommen werden.

Tabelle 90: RUMM 2030: Kompetenzstufen in Bezug auf den CareMaTex

Kompetenzstufe	Punkte	Personenparameter	Itemparameter
0	0	-4,036	-4,883
	1	-3,125	-3,972
	2	-2,415	-3,262
	3	-1,864	-2,711
Schwelle der Kompetenzstufe 0 zu I			
I	4	-1,382	-2,229
	5	-0,942	-1,789
	6	-0,531	-1,378
	7	-0,138	-0,985
	8	0,242	-0,605
Schwelle der Kompetenzstufe I zu II			
II	9	0,619	-0,228
	10	1,000	0,153
Schwelle der Kompetenzstufe II zu III			
III	11	1,399	0,552
	12	1,835	0,988
	13	2,345	1,498
	14	3,022	2,175

In die Kompetenzstufe 0 werden alle Probanden mit einem Gesamtpunktscore von 0 bis einschließlich 3 Punkten eingeordnet. In die erste Kompetenzstufe werden alle Probanden mit einer Gesamtpunktescore von 4 Punkten bis einschließlich 8 Punkten eingeteilt. Bei Kompetenzstufe II sind es die Probanden mit einem Gesamtpunktescore von 9 Punkten bis einschließlich 10 Punkten und bei Kompetenzstufe III sind es Probanden mit einem Gesamtpunktescore von 11 Punkten bis 14 Punkten. Anhand von Tabelle 91 lassen sich die Schülerfähigkeiten aufzeigen.

Tabelle 91: RUMM 2030: Schülerfähigkeit in Bezug auf den CareMaTex

	Hauptschulabschluss	Realschulabschluss	Fachhochschulabschluss/Abitur
0	14	13	1
1	12	13	5
2	20	12	2
3	14	14	0
4	5	10	5
5	10	6	6
6	8	11	1
7	6	17	2
8	3	2	3
9	2	6	3
10	0	1	4
11	0	0	2
12	0	0	2
13	0	1	1
14	0	1	1
15	0	0	0
	n = 94	n = 107	n = 38

Ermittelt wurde, dass 63,9 % der Hauptschüler die Kompetenzstufe 0, 34,0 % der Hauptschüler die Stufe I, 2,1 % der Hauptschüler die Kompetenzstufe II und 0,0 % der Hauptschüler die Kompetenzstufe III erreicht haben. Die Realschüler schneiden wieder vergleichsweise besser ab. Etwa 48,6 % der Realschüler erreichen die Kompetenzstufe 0 und 43 % die Stufe I, 6,5 % die Stufe II und 1,9 % die Stufe III. Im Vergleich zu den Hauptschülern und Realschülern schneiden die Schüler mit Fachhochschulreife beziehungsweise Fachabitur oder Abitur auch wieder besser ab. Von diesen Schülern erreichten 21,1 % die Kompetenzstufe 0 und 44,7 % die Stufe I, 18,4 % die Kompetenzstufe II und 15,8 % die Stufe III. Aus den Ergebnissen der Analyse mit der Software RUMM 2030 wird ersichtlich, dass sich auch hier große Unterschiede in Bezug auf das Kompetenzniveau zwischen den Schulabschlüssen ergeben, was in Tabelle 92 dargestellt wird.

Tabelle 92: Übersicht Kompetenzniveau sortiert nach Schulabschlüssen - Teil 5

Kompetenzstufen	Hauptschulabschluss	Realschulabschluss	Fachhochschulreife/ Abitur
Kompetenzstufe 0	63,9 %	48,6 %	21,1 %
Kompetenzstufe I	34,0 %	43,0 %	44,7 %
Kompetenzstufe II	2,1 %	6,5 %	18,4 %
Kompetenzstufe III	0,0 %	1,9 %	15,8 %

Diskussion der Ergebnisse aus der Analyse mit eRm: Nach der Elimination der Items 9 und 12 werden im Folgenden die theoretisch erreichbaren Gesamtscores, aufgrund der Rasch-Analyse mit eRm, in eine der vier Kompetenzstufen eingeordnet. Tabelle 93 kann diese Zuordnung entnommen werden.

Tabelle 93: eRm: Kompetenzstufen in Bezug auf den CareMaTex

Kompetenzstufe	Punkte	Personenparameter	Itemparameter
0	0	-4.305	-5,152
	1	-3.367	-4,214
	2	-2.496	-3,343
	3	-1.900	-2,747
Schwelle der Kompetenzstufe 0 zu I			
I	4	-1.410	-2,257
	5	-0.975	-1,822
	6	-0.569	-1,416
	7	-0.180	-1,027
	8	0.203	-0,644
Schwelle der Kompetenzstufe I zu II			
II	9	0.589	-0,258
	10	0.989	0,142
Schwelle der Kompetenzstufe II zu III			
III	11	1.415	0,568
	12	1.894	1,047
	13	2.479	1,632
	14	3.342	2,495
	15	4.272	3,425

In die Kompetenzstufe 0 werden alle Probanden mit einem Gesamtpunktscore von 0 bis einschließlich 3 Punkten eingeordnet. In die erste Kompetenzstufe werden alle Probanden mit einem Gesamtpunktscore von 4 Punkten bis einschließlich 8 Punkten eingeteilt. Bei Kompetenzstufe II sind es die Probanden mit einem Gesamtpunktscore von 9 Punkten bis einschließlich 10 Punkten und bei Kompetenzstufe III sind es Probanden mit einem Gesamtpunktscore von 11 Punkten bis 15 Punkten. Anhand von

Tabelle 94 lassen sich die Schülerfähigkeiten in Form von Kompetenzstufen einteilen.

Tabelle 94: eRm: Schülerfähigkeiten in Bezug auf den CareMaTex

	Hauptschulabschluss	Realschulabschluss	Fachhochschulabschluss/Abitur
0	13	13	1
1	13	12	3
2	16	10	4
3	14	13	0
4	10	12	5
5	8	6	5
6	6	9	2
7	7	15	1
8	2	5	2
9	3	8	3
10	2	1	3
11	0	1	3
12	0	0	2
13	0	1	2
14	0	1	2
15	0	0	0
	n = 94	n = 107	n = 38

Es ergibt sich, dass 59,6 % der Hauptschüler die Kompetenzstufe 0, 35,1 % die Stufe I, 5,3 % die Kompetenzstufe II und 0,0 % die Kompetenzstufe III erreicht haben. Die Realschüler schneiden erneut vergleichsweise besser ab. Etwa 44,9 % von ihnen erreichen die Kompetenzstufe 0 und 43,9 % die Stufe I, 8,4 % die Stufe II und 2,8 % die Stufe III. Im Vergleich zu den Hauptschülern und Realschülern schneiden die Schüler mit Fachhochschulreife beziehungsweise Fachabitur oder Abitur auch erneut besser ab. Denn 21,0 % der Schüler erreichen die Kompetenzstufe 0 und 39,5 % die Stufe I, 15,8 % die Kompetenzstufe II und 23,7 % die Stufe III. Aus den Ergebnissen der Analyse mit eRm wird deutlich, dass sich wiederholt große Unterschiede in Bezug auf das Kompetenzniveau zwischen den Schulabschlüssen ergeben, was Tabelle 95 dargestellt.

Tabelle 95: Übersicht Kompetenzniveau sortiert nach Schulabschlüssen - Teil 6

Kompetenzstufen	Hauptschulabschluss	Realschulabschluss	Fachhochschulreife/ Abitur
Kompetenzstufe 0	59,6 %	44,9 %	21,0 %
Kompetenzstufe I	35,1 %	43,9 %	39,5 %
Kompetenzstufe II	5,3 %	8,4 %	15,8 %
Kompetenzstufe III	0,0 %	2,8 %	23,7 %

Zusammenfassung: Die Ergebnisse zeigen, dass die untersuchten Altenpflegeschüler nicht besser abschnitten als in der Untersuchung in Bezug auf den TeMaTex, obwohl die Testaufgaben des TeMaTex zuvor an den Kontext der Altenpflege angepasst wurden. Es wurde auch in dieser Erhebung deutlich, dass es in Bezug auf das mathematische Textverständnis große Unterschiede zwischen den Schulabschlüssen gibt. Im Rahmen dessen ist zu erwähnen, dass die Hauptschüler wieder am schlechtesten abgeschnitten haben.

12.3 Diskussion zu den Forschungsfragen und Hypothesen

Die erste Forschungsfrage und die daraus abgeleiteten Hypothesen beziehen sich auf die inhaltsvalide Darstellung des mathematischen Textverständnisses mithilfe der Items aus dem TeMaTex (vgl. Jordan 2011a). Die 18 Aufgaben des TeMaTex behandeln folgende mathematischen Bereiche: *Grundrechenarten (Arithmetik), Prozentrechnung, Dreisatz, Bruchrechnung und Gleichungen*. Im Rahmen der Analysen zur Erstellung eines mathematischen Anforderungsprofils für den Altenpflegeberuf in Kapitel 3 wurde deutlich, dass zumindest einer der genannten mathematischen Bereiche – der Bereich *Grundrechenarten* – eine hohe Relevanz für die Altenpflegeausbildung aufweist. Mit den Items 1 bis 6 des TeMaTex werden folgende mathematische Anforderungen gestellt: Flächenberechnung, Umfangberechnung, Berechnung der Mantelfläche, Arithmetik (Multiplikation, Addition, Subtraktion) und Prozentrechnung (Prozentwert). Mit Item 7 bis 9 wird der Bereich der Arithmetik (Addition, Multiplikation, Division) erfasst. Item 10 bis 13 können den Bereichen Arithmetik (Addition, Verdopplung), Bruchrechnung und Prozentrechnung (Prozentzahl) zugeordnet werden. Item 14 bis

18 umfassen die Bereiche Arithmetik (Multiplikation, Division, Subtraktion) und Gleichungen (Satz des Pythagoras).

Mithilfe des TeMaTex wird versucht, das Konstrukt des mathematischen Textverständnisses zu quantifizieren. Die Einteilung in Kompetenzstufen über die Summenscores der Probanden würde eine Einschätzung erlauben, mit welchem Niveau die Altenpflegeschüler ihre Ausbildung beginnen. Damit dies gewährleistet werden kann, sollten alle Dimensionen, die dem Konstrukt des mathematischen Textverständnisses zugeschrieben werden, durch die Items des TeMaTex erfasst werden. Im Rahmen der inhaltlichen Bewertung der Aufgaben des TeMaTex wurde festgestellt, dass alle wichtigen mathematischen Dimensionen des mathematischen Textverständnisses vertreten sind. Dies bedeutet, dass alle für den Altenpflegeberuf relevanten mathematischen Anforderungen durch die Aufgaben des TeMaTex abgedeckt werden. Anzumerken ist allerdings, dass die mathematischen Bereiche wie Flächenberechnung, Umfangberechnung, Berechnung der Mantelfläche oder der Satz des Pythagoras nicht zu den primären mathematischen Anforderungsbereichen in der Altenpflege gehört. Des Weiteren ist zu konstatieren, dass die Itemschwierigkeiten und die Personenfähigkeiten der Altenpflegeschüler nicht zueinander passen (siehe Abbildung 9). Es zeigte sich, dass die Personenfähigkeiten im Vergleich zu den Itemschwierigkeiten geringer sind, das heißt, dass die Items des TeMaTex zu schwer sind.

Vor diesem Hintergrund muss die H_0 -Hypothese verworfen und die H_1 -Hypothese angenommen werden, nämlich: „Die Items des Tests zur Erfassung des mathematischen Textverständnisses stellen das mathematische Textverständnis von Altenpflegeschülern zu Beginn der Ausbildung nicht inhaltsvalide dar. Sie bilden die latente Variable des mathematischen Textverständnisses nicht ab“.

Die zweite Forschungsfrage und die daraus abgeleiteten Hypothesen versuchen zu eruieren, ob die dichotomen Bewertungen der Items des Te-

MaTex intervallskaliert behandelt, aufsummiert und in Notenwerte respektive transformiert werden dürfen. In Kapitel 5.4 wurde bereits der dichotome Charakter der Items beschrieben. Die Ergebnisse der Rasch-Analyse mit den Software-Programmen WINMIRA und eRm zeigten jeweils eine Modellgeltung (siehe Tabelle 40 und Tabelle 42). Dies bedeutet, dass damit einhergehend eine Transformation in intervallskalierte Daten nachgewiesen werden konnte. Die Analysen mit dem Programm RUMM 2030 führten jedoch zu abweichenden Ergebnissen. Um eine Modellgeltung erreichen zu können, mussten vier Items aus dem Test eliminiert werden (siehe Tabelle 32 und Tabelle 33).

Wie in Kapitel 12.1 beschrieben, arbeitet das Statistik-Programm RUMM 2030 restriktiver und liefert aussagekräftigere Ergebnisse hinsichtlich einer Modellgeltung. Daher dienen diese Ergebnisse als Grundlage zur Beantwortung der vorliegenden Forschungsfrage und als Hilfe zur Entscheidung, ob die zur Forschungsfrage formulierte Nullhypothese beibehalten werden kann oder verworfen werden muss. Daher ist die zweite Forschungsfrage zu verneinen und die H_0 -Hypothese zu verwerfen. Die H_1 -Hypothese: „Die dichotomen Items des TeMaTex können nicht intervallskaliert behandelt werden. Einzelne Bewertungen von Altenpflegeschülern dürfen nicht aufsummiert und in Notenwerte respektive Kompetenzstufen transformiert werden.“ wird beibehalten.

Die dritte Forschungsfrage und deren Hypothesen beinhalten die konstruktvalide Abbildung des mathematischen Textverständnisses mit den Aufgaben des TeMaTex. Basierend auf den dargelegten Ergebnissen konnte nur in Abhängigkeit vom gewählten Software-Programm die Konstruktvalidität mit dem Rasch-Modell validiert werden. Laut dem Wald-Test, welcher mit eRm durchgeführt wurde, liegt eine Itemheterogenität vor (siehe Tabelle 43). Darüber hinaus wurde mittels des grafischen Modelltests eine Personheterogenität festgestellt (siehe Abbildung 13).

Aufgrund der vorliegenden Ergebnisse wird die H_0 -Hypothese verworfen und die H_1 -Hypothese angenommen: „Nach der in dieser Arbeit durchgeführten Untersuchung bilden die Items des TeMaTex das mathematische Textverständnis von Altenpflegeschülern zu Ausbildungsbeginn nicht konstruktvalid ab.“

Die vierte Forschungsfrage bezog sich auf das Niveau des mathematischen Textverständnisses, über welches die Altenpflegeschüler zu Beginn der Pflegeausbildung verfügen. Vor dem Hintergrund der Rasch-Analysen mit dem Statistikprogramm RUMM 2030 dürfen für die 18 Items des TeMaTex keine Summenwerte gebildet werden beziehungsweise Einteilungen in Kompetenzstufen und daraus Ableitungen zum Niveau der Altenpflegeschüler hinsichtlich des mathematischen Textverständnisses zu Beginn der Ausbildung erfolgen. Aufgrund der Tatsache, dass die Ergebnisse der Rasch-Analysen mit der Statistik-Software WINMIRA eine Modellgeltung aufzeigten, konnten die vorliegenden Summenwerte interpretiert werden und mit den Ergebnissen aus der Studie von Jordan (2011) verglichen werden (siehe Kapitel 12.1), da die Ergebnisse von Jordan auch mit dem Statistik-Programm WINMIRA erzeugt wurden. Die Ergebnisse der vorliegenden Studie zeigten auf, dass die Hauptschüler und Realschüler im Vergleich zu den Schülern mit Fachhochschulreife respektive Abitur ein geringeres Niveau des mathematischen Textverständnisses zu Ausbildungsbeginn besitzen (siehe Tabelle 80).

Die fünfte Forschungsfrage beschäftigte sich damit, ob es möglich ist, die Items des TeMaTex an den Kontext der Altenpflege anzupassen, um ein konstruktvalides Testinstrument zur Erfassung des mathematischen Textverständnisses in der Altenpflege (CareMaTex) zu konstruieren. Die Analysen in Kapitel 9 und 10 haben gezeigt, dass der CareMaTex eine Inhalts- und Konstruktvalidität aufweist. Obwohl in Kapitel 9 keine statistisch signifikante Urteilsübereinstimmung der Experten zustande kam und daraus schlussfolgernd alle 17 Items aus dem Expertenrating im Itemsatz belassen werden könnten, stellt sich die Frage der inhaltlichen Validität von Aufgabe

16. Aufgabe 16 umfasst eine mathematische Operation (Satz des Pythagoras), die für die Altenpflege keine Relevanz aufzeigt (vgl. Kapitel 3 und 9). Im Rahmen der Analysen mit RUMM 2030 mussten Item 14 und 16 eliminiert werden, damit eine Modellgeltung zustande kam. Aufgrund der Tatsache, dass Item 16 inhaltlich und statistisch große Defizite aufweist, wird es wie Item 14 in der Endversion nicht berücksichtigt. Die Analysen mit WINMIRA und RUMM 2030 ergaben, dass nach dieser Itemelimination eine Modellgeltung zustande kam. Die Endversion des CareMaTex umfasst somit 15 Items.

Das Ziel der Untersuchung war es unter anderem, ein konstruktvalides Testinstrument zur Erfassung des mathematischen Textverständnisses von Altenpflegeschülern zu Beginn der Pflegeausbildung zu erhalten, mit dem valide Summenwerte aus dichotomen Daten zuverlässig in Kompetenzstufen transformiert werden können. Dieses Ziel wurde in der vorliegenden Untersuchung für den TeMaTex nicht erreicht. Für den TeMaTex war es möglich, für 14 von insgesamt 18 Aufgaben die Raschvalidität nachzuweisen. Raschvalide Items sind stärker in ihrer Differenzierungsleistung und unterscheiden die Personen in Bezug auf das mathematische Textverständnis besser. Für diese Items ist eine Transformation in Kompetenzstufen erlaubt.

Die jetzige Version des TeMaTex ist, wie bereits ausgeführt, aufgrund der fehlenden inhaltlichen Validität, der methodischen Probleme sowie der daraus resultierenden falschen Schlussfolgerungen zur Erfassung des mathematischen Textverständnisses von Altenpflegeschülern zu Beginn der Pflegeausbildung ungeeignet.

13 Fazit und Ausblick

Ein niedriges Niveau des mathematischen Textverständnisses zu Beginn der Altenpflegeausbildung steht im direkten Widerspruch zur gesetzlichen Forderung, die Altenpflege auf einem hohen fachlichen und wissenschaftlichen Niveau anzubieten. Dies kann kaum gelingen, wenn die notwendigen Fähigkeiten der Auszubildenden im Bereich des mathematischen Textverständnisses fehlen und diese im Rahmen der Ausbildung auch nicht gebildet oder gefördert werden können, weil hierfür keine Lehrinhalte vorgesehen sind. Es lässt sich vermuten, dass durch die Novellierung des Altenpflegegesetzes und die sich daraus ableitenden Maßnahmen – zum Beispiel die Durchführung von Kompetenzfeststellungsverfahren, wobei im Rahmen dessen lediglich die Dimensionen der Handlungskompetenz (Fach,- Sozial,- Methoden- und personelle Kompetenz) geprüft werden – diesem Trend nicht entgegengewirkt werden kann. Vielmehr wird hierbei erneut versucht, mehr „fähiges“ Personal zu rekrutieren, um dem Personal-mangel in der Altenpflege zu begegnen. Dieser Umstand lässt vermuten, dass es für das Erlernen des Berufes der Altenpflege keiner Fähigkeiten oder nur geringfügiger Fähigkeiten in Bezug auf das mathematische Textverständnis bedarf. Berufspolitisch gesehen ist diese Situation äußerst bedenklich und könnte sich deprofessionalisierend auf den Altenpflegeberuf auswirken. Arbeitsmarktpolitisch ist es natürlich für die entsprechenden Akteure von großer Bedeutung, dem Fachpersonalmangel in der Altenpflege zu begegnen. Das Postulat der Politik *Jeder kann pflegen!* darf nicht als Grundlage berufspolitischer Aktionen gelten. Das Problem liegt bereits in der Semantik der Begrifflichkeit *Fachpersonalmangel*. Es ist durchaus denkbar und auch wichtig, ungelerntes Personal wie Ehrenamtliche in die direkte Pflege einzubeziehen. Es ist aber ein großer Fehler, davon auszugehen, dass durch das umfangreiche Rekrutieren von potentiell ungeeigneten Personen dem Fachkräftemangel entgegengewirkt werden kann. Natürlich ist es so, dass in erster Linie quantitativ dem Mangel Abhilfe geschaffen wird. Jedoch ist es von zentraler Bedeutung für die Altenpflegerische Arbeit und nicht zuletzt für die Bewohner und Klienten, dass sie in erster Linie von

qualifiziertem Pflegefachpersonal betreut werden. Diese Fachlichkeit wird erreicht, in dem diese Menschen mindestens eine dreijährige Altenpflegeausbildung absolviert und das Ausbildungsziel erreicht haben, sodass sie eigenverantwortlich und selbstständig handeln können und darüber hinaus eine ausgeprägte Handlungskompetenz aufweisen. Hierbei ist die Förderung beziehungsweise vereinzelt sicher auch die grundständige Ausbildung von Kulturtechniken wichtig. Dazu gehört die Lesefähigkeit, die Rechenfähigkeit (insbesondere das mathematische Textverständnis) und die Sprachfähigkeit. Ein Blick in die alltägliche Pflegepraxis verrät, dass es unerlässlich ist, ein hohes Niveau im Bereich des mathematischen Textverständnisses zu besitzen. In der Praxis gibt es immer wieder Situationen, in denen etwas berechnet werden muss (Zubereitung von Medikamenten, Tropfenzahl bzw. Tropfgeschwindigkeit einer Infusionslösung). Es müssen Arztbriefe, Pflegeberichte oder Fachartikel gelesen und verstanden werden. Vor diesem Hintergrund ist es zukünftig von zentraler Bedeutung, die Kulturtechniken als eine Art Querschnittsdimensionen der beruflichen Handlungskompetenz zu betrachten, welche alle anderen (Sub-) Dimensionen wie Fach-, Methoden-, Sozial- und Personalkompetenz direkt beeinflussen.

In Bezug auf Empfehlungen für weitere Arbeiten in der Pflegeforschung sei in erster Linie darauf hingewiesen, dass es notwendig ist, Instrumente zur Erfassung des mathematischen Textverständnisses für alle Pflegeberufe zu entwickeln. Die Ergebnisse dieser Untersuchung haben ergeben, dass die Anwendung des TeMaTex für eine Einzelfalldiagnostik im Rahmen von Personalauswahlverfahren oder im Rahmen der Ausbildung nicht zu empfehlen ist. Die vorliegende Arbeit gibt Impulse dafür (vgl. Kapitel 11), wie eine solche Entwicklung von Testaufgaben zur Erfassung des mathematischen Textverständnisses für alle Pflegeberufe aussehen könnte. Von zentraler Bedeutung ist hierbei der Bezug auf ein theoretisches Gerüst, welches zur Entwicklung eines aussagekräftigen Testinstrumentes notwendig ist. Als theoretisches Gerüst ist hier das Konstrukt der *quantitative Literacy* zu erwähnen, welches bereits in der IALS-Studie und bei Jordan (2011) als Grundlage zur Erfassung des mathematischen Textverständnisses diente.

Darauf aufbauend ist es wichtig, relevante Dimensionen des Konstruktes aus tiefgreifenden Analysen in den Bereichen *exchange* und *use value* zu gewinnen. Dies bedeutet, dass zuerst die bestehenden Curricula zur Altenpflege sowie alle Lehrwerke in Bezug auf mathematische Inhalte gesichtet und analysiert werden müssen. Danach sollten die gewonnenen Dimensionen auf ihre Altenpflegepraktische Relevanz überprüft werden. Dies müsste über eine Expertenbefragung erfolgen. In Kapitel 3 wurde ersichtlich, dass in der theoretischen Altenpflegeausbildung lediglich der mathematische Bereich *Grundrechenarten* fokussiert wird. Dies ist deutlich zu kurz gegriffen. In der alltäglich Altenpflegerischen Arbeit begegnen den Auszubildenden deutlich vielfältigere mathematische Anforderungen als das Berechnen der Tropfengeschwindigkeit einer Infusionslösung, wofür das Anwenden eines Dreisatzes beziehungsweise der Prozentrechnung wichtig ist.

Die Arbeit hat aufgezeigt, mit welchem Niveau die Altenpflegeschüler bezüglich des mathematischen Textverständnisses ihre Ausbildung beginnen. Die Ergebnisse sind bedenklich. Es wurde deutlich, dass die Hauptschüler am schlechtesten abschnitten. Dies kann natürlich daran liegen, dass die Ergebnisse aufgrund der zu geringen Stichprobengrößen zustande kamen. Daraus lässt sich ein weiterer Forschungsbedarf benennen: Es ist offensichtlich notwendig, das mathematische Textverständnis von Altenpflegeschülern in einer größeren Stichprobe zu erfassen. Hierfür müsste allerdings, wie bereits erwähnt, zuerst ein entsprechendes Instrument zur Einzelfalldiagnostik entwickelt werden. Jedoch lässt sich aus den Erkenntnissen vorliegender Untersuchung eine bedenkliche Tendenz ableiten. Mit einer niedrigen Zugangsvoraussetzung geht auch ein niedriges Niveau des mathematischen Textverständnisses der Auszubildenden und vermutlich auch ein niedriges Niveau in Bezug auf alle anderen Kulturtechniken einher. Im Gegenzug schnitten die Altenpflegeschüler, welche über Fachhochschulreife oder Abitur besitzen, erheblich besser ab als die Hauptschüler. Es sei festgehalten, dass die Diagnostik solcher Defizite in der Berufsausbildung nicht allein zur Lösung der Problematik beiträgt. Es müssen daran anschließend Fördermaßnahmen entwickelt werden, die dieser Problematik

entgegenwirken können. Dies konnte hier nicht bewerkstelligt werden und wird daher als weiterer Forschungsbedarf festgestellt, wobei auch auf ausbildungsflankierende Förderprogramme, wie zum Beispiel auf Nachhilfeunterricht zu medizinischem Rechnen, verwiesen sei, welche in die bestehenden Schulcurricula integriert werden müssten.

Die noch zu entwickelnden Testinstrumente zur Erfassung des mathematischen Textverständnisses für alle Pflegeberufe könnten in Zukunft im Rahmen von Bewerberauswahlverfahren beziehungsweise Kompetenzfeststellungsverfahren sowie für die Einzelfalldiagnostik im Rahmen des Unterrichtes und in der Pflegeausbildung angewendet werden. Es wäre allerdings zu kurz gegriffen, wenn zur Ermittlung der Ausbildungsfähigkeit nur das mathematische Textverständnis berücksichtigt würde. Vielmehr gehören die Lesefähigkeit und die Sprachfähigkeit dazu.

14 Literatur

- Andrich, D.; Sheridan, B.; Luo, G. (2009): Rasch Unidimensional Measurement Models (RUMM). Interpreting RUMM 2030 Part I Dichotomous Data. Version 2030: Rumm Laboratory Pty Ltd.
- Artelt, C.; Stanat, P.; Schneider, W.; Schiefele, U. (2001): Lesekompetenz. Testkonzeption und Ergebnisse. In: E. Klieme M.; Neubrand M.; Prenzel U.; Schiefele W.; Baumert, E. (Hg.): PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Opladen: Leske u. Budrich, S. 69-137.
- Becker-Mrotzek, M.; Kusch, E. (2007): Sachtexte lesen und verstehen. In: Der Deutschunterricht 59 (1), S. 31-38.
- Blumenauer, H. (2013): Kompetenzfeststellung im Rahmen eines Gesetzes zur Stärkung der beruflichen Aus- und Weiterbildung in der Altenpflege. Online verfügbar unter http://www.perspektive-berufsabschluss.de/downloads/Downloads_Projekte_Nachqualifizierung/Nachqualifizierung_Altenpflege_Fachartikel_27_02_2013.pdf, zuletzt geprüft am 06.09.2013.
- Bortz, J.; Lienert, G. A. (2003): Kurzgefasste Statistik für die klinische Forschung. Leitfaden für die verteilungsfreie Analyse kleiner Stichproben ; mit 91 Tabellen. 2. Aufl. Heidelberg: Springer (Springer-Lehrbuch).
- Brühl A. (2012): Methodologische Orientierung der Pflegewissenschaft bei der Entwicklung standardisierter Messinstrumente. In: Brühl A. (2012): Pflegebedürftigkeit messen? Herausforderungen bei der Entwicklung pflegerischer Messinstrumente am Beispiel des Neuen Begutachtungsassessments (NBA). Wissenschaftlicher Bericht des Lehrstuhls für Statistik und standardisierte Verfahren der Pflegeforschung, online verfügbar unter: https://kidoks.bsz-bw.de/files/30/Bruehl_Pflegebeduerftigkeit_messen_2012.pdf, zuletzt geprüft am 07.07.2016.
- Bühner, M. (2006): Einführung in die Test- und Fragebogenkonstruktion. 2. Aufl. München [u.a.]: Pearson Studium (PS Psychologie).
- Bühner, M. (2011): Einführung in die Test- und Fragebogenkonstruktion. 3. Aufl. München [u.a.]: Pearson Studium (PS Psychologie).
- Bundesagentur für Arbeit (2016): BERUFENET Steckbrief – Altenpfleger/in. Online verfügbar unter <https://berufenet.arbeitsagentur.de/berufenet/bkb/9065.pdf>, zuletzt geprüft am 24.02.2016.
- Bundesregierung der Bundesrepublik Deutschland (2002): Ausbildungs- und Prüfungsverordnung für den Beruf der Altenpflegerin und des Altenpflegers (Altenpflege-Ausbildungs- und Prüfungsverordnung – AltPflAPrV). AltPflAPrV, vom 06.12.2011. Online verfügbar unter <http://www.gesetze-im-internet.de/bundesrecht/altpflaprv/gesamt.pdf>, zuletzt geprüft am 06.09.2013.
- Bundesregierung der Bundesrepublik Deutschland (2003): Gesetz über die Berufe in der Altenpflege (Altenpflegegesetz- AltPflG). AltPflG, vom

13.03.2013. Online verfügbar unter <http://www.gesetze-im-internet.de/bundesrecht/altpflg/gesamt.pdf>, zuletzt geprüft am 06.09.2013.

Bundesregierung der Bundesrepublik Deutschland (2013): Gesetz zur Stärkung der beruflichen Aus- und Weiterbildung in der Altenpflege, vom 18.03.2013. Fundstelle: BGBl. Teil I Nr. 13, S. 446 - 447. Online verfügbar unter <http://npl.ly.gov.tw/pdf/8184.pdf>, zuletzt geprüft am 17.05.2013.

Coben, D. (2002): Use Value and Exchange Value in Discursive Domains of Adult Numeracy Teaching. *LITERACY AND NUMERACY STUDIES*, Jg. 11, Nr. 2, S. 25-35.

Derrerr-Merk, Elfriede (2013): Prüfungswissen Altenpflege. Vorbereitung kompakt - alle prüfungsrelevanten Lernfelder. 2. Aufl. München: Urban & Fischer in Elsevier.

Deutsches Institut für Internationale Pädagogische Forschung (DIPF) (2012): Bildung in Deutschland 2012. Ein indikatorengestützter Bericht mit einer Analyse zur kulturellen Bildung im Lebenslauf. Bielefeld. Online verfügbar unter http://www.bildungsbericht.de/daten2012/bb_2012.pdf, zuletzt geprüft am 25.04.2013.

Dijk, T. A. v. & Kintsch, W. (1983): *Strategies of Discourse Comprehension*. New York: Academic Press.

Eid, M.; Schmidt, K. (2014): *Testtheorie und Testkonstruktion*. Göttingen: Hogrefe (Bachelorstudium Psychologie).

Fischer, G. H.; Molenaar, I. W. (1995): *Rasch Models. Foundations, recent developments, and applications ; [papers originally presented at a workshop held at the University of Vienna, Feb. 25 - 27, 1993]*. New York: Springer.

Fricke, A.; Besuden, H. (1987): *Sachrechnen. Das Lösen angewandter Aufgaben*. 1. Aufl. Stuttgart: Klett.

Gartz, M.; Hüchtermann, M.; Mrytz, B. (1999): *Schulabgänger. Was sie können und was sie können müssten*. Köln: Deutscher Institutsverl (Kölner Texte und Thesen. 53).

Gierse, M. (2001): *Fachrechnen für Pflegeberufe: Grundrechenarten und medizinisches Fachrechnen*. Hannover: Schlütersche.

Grundmann, H. (2007): *Sprachfähigkeit und Ausbildungsfähigkeit. Der berufsschulische Unterricht vor neuen Herausforderungen*. Baltmannsweiler: Schneider Verlag Hohengehren.

Greefrath, G. (2010). *Didaktik des Sachrechnens in der Sekundarstufe*. Heidelberg: Spektrum.

Hambleton, R. K.; Swaminathan, H.; Rogers, H. J. (1991): *Fundamentals of item response theory*. 1. print. Newbury Park, Calif.: Sage Publ. (Measurement methods for the social sciences series, 2).

Heene M., Draxler C., Ziegler M., Bühner M. (2011): Performance of the bootstrap Rasch model test under violations of non-intersecting item response functions. In: Psychological Test and Assessment Modeling, Vol. 53 (3), S. 283-294

Hessisches Sozialministerium (Hg.) (2013): Rahmenlehrplan für die schulische und betriebliche Ausbildung Fachkraft Altenpflege. Online verfügbar unter https://soziales.hessen.de/sites/default/files/media/hsm/rahmenlehrplan_altenpflege_2011.pdf, zuletzt geprüft am 07.07.2016.

Ingenkamp, K., & Lissmann, U. (1997): Lehrbuch der pädagogischen Diagnostik. Weinheim/Basel: Beltz Verlag.

International Adult Literacy Survey (IALS) (2000): Literacy in the information age. Final report of the International Adult Literacy Survey. Paris, [Ottawa]: Organisation for Economic Cooperation and Development; Statistics Canada.

Jordan, R. (2011a): Entwicklung und Validierung eines Testverfahrens zur Ermittlung der Lesekompetenz und des mathematischen Textverständnisses mit empirischer Untersuchung an allgemeinbildenden und berufsbildenden Schulen. Münster: WTM, Verl. für wiss. Texte und Medien (Testentwicklung und Evaluation in der Mathematik-Didaktik. 1).

Jordan, R.; Stein, M. (2011b): TeMaTex. Online verfügbar unter <https://www.testzentrale.de/shop/test-zum-mathematischen-textverstaendnis.html>, zuletzt geprüft am 07.07.2016.

Jordan, R.; Stein, M. (2011c): TeMaTex -Test zum mathematischen Textverständnis. Ein Testverfahren zur Ermittlung des mathematischen Textverständnisses am Ende der Sekundarstufe I bzw. zu Beginn der beruflichen Ausbildung. Münster: WTM, Verl. für wiss. Texte und Medien (Testentwicklung und Evaluation in der Mathematik-Didaktik. 1).

Kintsch, W. (1988): The role of knowledge in discourse comprehension. A construction-integration model. Psychological Review, 95, 163-182.

Kintsch, W. (1994): Kognitionspsychologische Modelle des Textverstehens: Literarische Texte. In Reusser, K., & Reusser-Weyeneth, M., Verstehen. Psychologischer Prozess und didaktische Aufgabe (S. 39-53). Bern: Huber.

Klieme, E.; Baumert, J.; Köller, O.; Bos, W. (2000): Mathematische und naturwissenschaftliche Grundbildung. Konzeptuelle Grundlagen und die Erfassung und Skalierung von Kompetenzen. In: TIMSS/III: Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn ; Vol. 1. Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit.

Knappich, T. (2012): Messung allgemeiner Kompetenzen von Altenpflegern mit dem TIMSS. Masterarbeit. Philosophisch-Theologische Hochschule, Vallendar. Pflegewissenschaftliche Fakultät.

Krauthausen, G.; Scherer, P. (2010): Einführung in die Mathematikdidaktik. 3. Aufl. Heidelberg: Spektrum Akad. Verl (Mathematik Primar- und Sekundarstufe).

Krupp, E. (2016): Die Pflege-Transparenzvereinbarung (PTVA) - ein Instrument zum Erfassen von Pflegequalität? Dissertation. Philosophisch-Theologische Hochschule, Vallendar. Pflegewissenschaftliche Fakultät. Online verfügbar unter: https://kidoks.bsz-bw.de/files/699/Krupp_E_Dissertation_2015.pdf

Leuders (2007): Kernlehrplan für das Gymnasium – Sekundarstufe I (G 8) in Nordrhein-Westfalen. Mathematik. Online verfügbar unter http://www.schulentwicklung.nrw.de/lehrplaene/upload/lehrplaene_download/gymnasium_g8/gym8_mathematik.pdf, zuletzt aktualisiert am 19.07.2007, zuletzt geprüft am 07.07.2016.

LoBiondo-Wood, G.; Haber, J. (1996): Pflegeforschung. Methoden, kritische Einschätzung und Anwendung. Berlin: Ullstein Mosby.

Maier, H. (1976): Didaktik der Mathematik 1 - 9. Sachlogische Analyse der Lehr- und Lernziele für die ersten neun Jahrgangsstufen. Donauwörth: Auer.

Mair, P.; Hatzinger, R. (2009a): Extended Rasch Modeling: The R Package eRm. PDF. Dateianhang zum Programmpaket eRm.

Mair, P.; Hatzinger, R. (2009b): Package 'eRm'. Extended Rasch Modeling. Version 0.10.2. Online verfügbar unter <http://r-forge.rproject.org/projects/erm>, zuletzt geprüft am 07.07.2016

Mitterhuber, D. (2008): Fachsprache und Lesekompetenz. In: Landesinstitut für Lehrerbildung und Schulentwicklung (Hg.): Förderung von Lesekompetenz – Unterrichtserfahrungen im Fachunterricht der Sekundarstufe I, 3. Teil der LI-Reihe Lesekompetenz. Hamburg, S. 19-23.

Moosbrugger, H.; Kelava, A. (Hg.) (2012a): Testtheorie und Fragebogenkonstruktion: Springer Berlin Heidelberg (Springer-Lehrbuch).

Moosbrugger, H.; Kelava, A. (2012b): Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien). In: Moosbrugger, H.; Kelava, A. (Hg.): Testtheorie und Fragebogenkonstruktion: Springer Berlin Heidelberg (Springer-Lehrbuch), S. 7-26.

Moosbrugger, H. (2012c): Klassische Testtheorie (KTT). In: Moosbrugger, H.; Kelava, A. (Hg.): Testtheorie und Fragebogenkonstruktion: Springer Berlin Heidelberg (Springer-Lehrbuch), S. 104-117.

Moosbrugger, H. (2012d): Item-Response-Theorie (IRT). In: Moosbrugger, H.; Kelava, A. (Hg.): Testtheorie und Fragebogenkonstruktion: Springer Berlin Heidelberg (Springer-Lehrbuch), S. 227-274.

OECD (2000): Schülerleistungen im internationalen Vergleich – Eine Rahmenkonzeption für die Erfassung von Wissen und Fähigkeiten. Berlin: Max-Planck-Institut für Bildungsforschung.

OECD (2005): Definition und Auswahl von Schlüsselkompetenzen. Zusammenfassung. Online verfügbar unter <http://www.oecd.org/pisa/35693281.pdf>, zuletzt geprüft am 06.09.2013.

Pospeschill, M. (2010): Testtheorie, Testkonstruktion, Testevaluation. München: Ernst Reinhard Verlag

Pospiech, U.; Bitterlich, A. (2007): "Alle wollen sie es schriftlich!" Formen und Funktionen des Schreibens im Beruf. In: Der Deutschunterricht 59 (1), S. 19-30.

Rasch, Björn; Frieze, Malte; Hofmann, Wilhelm; Naumann, Ewald (2006): Quantitative Methoden. Einführung in die Statistik. 2., erweiterte Auflage. Berlin Heidelberg: Springer Medizin Verlag Heidelberg (Springer-Lehrbuch).

Rasch, G. (1960): Probabilistic models for some intelligence and attainment tests. Copenhagen: Nielsen & Lydiche (Studies in mathematical psychology).

Rost, J. (2004): Lehrbuch Testtheorie – Testkonstruktion. Bern; Göttingen [u.a.]: Huber.

Reusser, K. (1992): Kognitive Modellierung von Text - , Situations - und mathematischem Verständnis beim Lösen von Textaufgaben. In: Reiss, K., Reiss, M., & Spandl, H., *Maschinelles Lernen*. Berlin: Springer Verlag.

Reusser, K. (1997). Erwerb mathematischer Kompetenzen: Literaturüberblick. In: Weinert, F. E.; Helmke, A. (Hg.), *Entwicklung im Grundschulalter* (S. 141-155). Weinheim: Beltz/Psychologie Verlags Union.

Schermelleh-Engel, K.; Werner, C. (2012): Methoden der Reliabilitätsbestimmung. In: Moosbrugger, H.; Kelava, A. (Hg.): *Testtheorie und Fragebogenkonstruktion*: Springer Berlin Heidelberg (Springer-Lehrbuch), S. 119-141.

Stein, M.; Winter, K. (2009): Das Projekt Mathe-Meister: Strukturen und Konzeption. Online verfügbar unter: https://www.mathematik.tu-dortmund.de/ieem/cms/media/BzMU/BzMU2009/Beitraege/STEIN_Martin_WINTER_Kathrin_2009_MatheMeister.pdf, zuletzt geprüft am 07.07.2016.

Stiftung Rechnen (Hg.) (2015): *Mathe4Job. Entdecke die Mathematik in deinem Beruf*. Unter Mitarbeit von Martin Stein. Münster: WTM Verlag für wissenschaftliche Texte und Medien.

Strobl, C. (2010): Das Rasch-Modell. Eine verständliche Einführung für Studium und Praxis. Mering: Rainer Hampp Verlag (Sozialwissenschaftliche Forschungsmethoden, 2).

UIE (1997): International Conference on Adult Education; 5th; Hamburger Deklaration zum Lernen im Erwachsenenalter: Agenda für die Zukunft; 1997. Online verfügbar unter <http://unesdoc.unesco.org/images/0011/001161/116114gero.pdf>, zuletzt aktualisiert am 31.01.2008, zuletzt geprüft am 07.07.2016.

Von Davier, M. (2001): Winmira 2001. Handbuch. Version 2001. Online verfügbar unter <http://208.76.84.140/~svfklumu/wmira/winmiramannual.pdf>, zuletzt geprüft am 07.07.2016.

Winter, H. (2003): Sachrechnen in der Grundschule. Problematik des Sachrechnens ; Funktionen des Sachrechnens ; Unterrichtsprojekte. 6. Aufl. Frankfurt am Main: Cornelsen Scriptor (Lehrer-Bücherei).

Wirtz, M.; Caspar, F. (2002): Beurteilerübereinstimmung und Beurteilerreliabilität. Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen. Göttingen: Hogrefe Verl. für Psychologie.

15 Verzeichnis der Abkürzungen

aktual.	aktualisierte
ALL	Adult Literacy and Lifeskills
AltPflG	Altenpflegegesetz
AltPflAPrV	Altenpflegeausbildungs- und Prüfungsverordnung
BMBF	Bildungsministerium für Bildung und Forschung
BMI	Body-Maß-Index
CareMaTex	Test zur Erfassung von mathematischem Textverständnis bei Altenpflegeschülern
CFT-R 20	Intelligenztest
cL	konditionale bzw. bedingte Likelihoodfunktion
cML	conditional bzw. bedingte Maximum Likelihood
df	Standardfehler
DIF	Differential Item Functioning
DIPF	Deutsches Institut für pädagogische Forschung
eRm	Abkürzung in Englisch für Softwarepackage extended Rasch modeling
ETS	Educational Testing Service
FIS	Fachinformationssystem Bildung
FitRes	Fit-Residual
Gl.	Gleichung
h	Stunde
IALS	International Adult Literacy Survey
ICC	Item Characteristic Curve
irr	Various Coefficients of Interrater Reliability and Agreement
IRT	Item Response Theory
KMK	Kultusministerkonferenz
KTt	Klassische Testtheorie
KW	Kalenderwoche
LTB-3	Lesetest
Locn	Logit-Werte
LR-Test	Likelihood-Ratio-Test
<i>m</i>	Rangreihen
ml	Milliliter
M	Mittelwert
MAR	missing-at-random
MC	Multiple-Choice
mm/Hg	Millimeter Quecksilbersäule

MLE	Maximum-Likelihood-Schätzung
n	Stichprobengröße
N	Stichprobengröße
OECD	Organisation for Economic Cooperation and Development
PISA	Programme for International Student Assessment
PTHV	Philosophisch-Theologische Hochschule Vallendar
1-PL Modell	1-Parameter Rasch-Modell
PTT	Probabilistische Testtheorie
R	eine freie Programmiersprache für statistische Berechnungen und Grafiken
r	Reliabilitätskoeffizient
®	Symbolabkürzung in Englisch für Registered
Rcmdr	R commander
RMSEA	Root Mean Square Error of Approximation
RUMM 2030	Rasch Unidimensional Measurement Models 2030
s.c.	subkutan
SD	Standardabweichung
SPS	Situation-Problem-Solver-Modell
SPSS	Abkürzung in Englisch für Statistical Package for the Social Science
TeMaTex	Test zur Erfassung des mathematischen Textverständnisses
TIMSS	Abkürzung in Englisch für Third International Mathematics and Science Study
uML	unbedingte Maximum Likelihood
UNESCO	United Nations Educational, Scientific and Cultural Organization
USA	United States of America
W	Konkordanzkoeffizient
WINMIRA	Standalone-Software-Paket für die Schätzung und Tests einer großen Anzahl von diskreten Mixture Modelle für kategoriale Variablen
WLE	Abkürzung in Englisch für Weighted Likelihood Estimates
WT	Wortschatztest
ZT	Zahlenfolgetest
ZVK	Zentraler Venenkatheter

16 Tabellenverzeichnis

Tabelle 1: Verfahren zur Kompetenzbilanzierung in Rheinland-Pfalz (nach Blumenauer 2013, S. 3 f.)	20
Tabelle 2: Lernbereiche (in Anlehnung an das Hessische Sozialministerium 2011, S. 11)	23
Tabelle 3: Lernfelder (in Anlehnung an das Hessische Sozialministerium 2011, S. 12 ff.)	24
Tabelle 4: Relevante Lernfelder	25
Tabelle 5: Analyse der Aufgaben 1 - 3 zur Fallsituation: Parenterale Ernährung	28
Tabelle 6: Infusionsgeschwindigkeit in ml/h.....	29
Tabelle 7: Aufgabe zur Prüfungsvorbereitung - Tabelle zur Aufgabenstellung (nach Derrer-Merk 2013, S. 161 f.)	33
Tabelle 8: Lösung zur Aufgabenstellung (nach Derrer-Merk 2013, S. 163)	35
Tabelle 9: Drei Bereiche der Literalität (nach OECD 2000, S. 10).....	39
Tabelle 10: Literalitätsstufen des Bereiches Quantitative Literacy (modifiziert nach Jordan 2011a, S. 96)	40
Tabelle 11: Charakteristika der jeweiligen Aufgabentypen (nach Maier 1976, S. 166)	43
Tabelle 12: Aufgabe 1 aus dem Beurteilungsbogen	55
Tabelle 13: Evaluationsfrage aus dem Beurteilungsbogen.....	56
Tabelle 14: Inhaltliche Beschreibung der Kompetenzstufen (nach Jordan 2011a, S. 351)	82
Tabelle 15: Raum- und Inventarplan der Stadt Karlsruhe (Erdgeschoss und 1. Obergeschoss) (nach Jordan & Stein 2011c, S. 3)	84
Tabelle 16: Aufgabenanalyse zu den Items 1 - 6 des Aufgabentextes 1 (nach Jordan 2011a, S. 199)	86
Tabelle 17: Aufgabenanalyse zu den Items 7 - 9 des Aufgabentextes 2 (nach Jordan 2011a, S. 201)	87
Tabelle 18: Aufgabenanalyse zu den Items 10 - 13 des Aufgabentextes 3 (nach Jordan 2011a, S. 203).....	88
Tabelle 19: Aufgabenanalyse zum Item 14 des Aufgabentextes 4 (nach Jordan 2011a, S. 205)	89
Tabelle 20: Aufgabenanalyse zu den Items 15+16 des Aufgabentextes 5 (nach Jordan 2011a, S. 205).....	89
Tabelle 21: Aufgabenanalyse zum Item 17 des Aufgabentextes 6 (nach Jordan 2011a, S. 206)	90
Tabelle 22: Aufgabenanalyse zum Item 18 des Aufgabentextes 7 (nach Jordan 2011a, S. 207)	90
Tabelle 23: Verteilung Schulabschlüsse - empirischer Teil I	95
Tabelle 24: Antworthäufigkeiten der 18 Items des TeMaTex - empirischer Teil I	96
Tabelle 25: Verteilung der Gesamtpunktwerte nach Schulabschluss - empirischer Teil I	97
Tabelle 26: Allgemeine Daten zur Rasch-Analyse - empirischer Teil I	98

Tabelle 27: Überblick Rasch-Analyse - empirischer Teil I	99
Tabelle 28: RUMM 2030: Überblick Rasch-Analyse mit RUMM 2030 - empirischer Teil I	99
Tabelle 29: RUMM 2030: Übersicht Item-Person und Item-Trait Interaction - empirischer Teil I.....	101
Tabelle 30: RUMM 2030: Übersicht Individual Item-Fit Werte - empirischer Teil I	102
Tabelle 31: RUMM 2030: Ergebnisse Fit-Residuen basierender Itemreduktion - empirischer Teil I.....	103
Tabelle 32: RUMM 2030: Ergebnisse χ^2 basierende Itemreduktion – empirischer Teil I	103
Tabelle 33: RUMM 2030: Übersicht individual Item-Fit Werte der vier eliminierten Items -empirischer Teil I	103
Tabelle 34: RUMM 2030: Zusammenfassung Differential Item Functioning der 18 Items des TeMaTex - empirischer Teil I.....	104
Tabelle 35: WINMIRA: Übersicht Personenparameter - empirischer Teil I.....	105
Tabelle 36: WINMIRA: Übersicht Mittelwert Personenparameterschätzung - empirischer Teil I	106
Tabelle 37: WINMIRA: Übersicht Itemschwierigkeiten der Items - empirischer Teil I	106
Tabelle 38: WINMIRA: Übersicht Q-Indizes und ZQ-Werte - empirischer Teil I	107
Tabelle 39: WINMIRA: Übersicht Ergebnisse Modellgeltungstest - empirischer Teil I	108
Tabelle 40: WINMIRA: Übersicht Ergebnisse Bootstrap - empirischer Teil I.....	109
Tabelle 41: eRm: Übersicht Ergebnisse Rasch-Analyse - empirischer Teil I.....	109
Tabelle 42: eRm: Ergebnisse Modellgeltungstest - empirischer Teil I....	110
Tabelle 43: eRm: Ergebnisse Wald-Test - empirischer Teil I.....	111
Tabelle 44: eRm: Ergebnisse Modellgeltungstest - empirischer Teil I....	112
Tabelle 45: eRm: Übersicht Personenparameter - empirischer Teil I	113
Tabelle 46: Übersicht Antwortverhalten.....	116
Tabelle 47: Auswertung Expertenrating.....	116
Tabelle 48: Lösungsweg zur Aufgabe 3	117
Tabelle 49: Lösungsweg zur Aufgabe 16	118
Tabelle 50: Ergebnisse des Expertenratings	119
Tabelle 51: Übersicht Ergebnisse der Berechnung Urteilsübereinstimmung	119
Tabelle 52: Verteilung Schulabschlüsse - empirischer Teil II	121
Tabelle 53: Antworthäufigkeiten der 17 Items des CareMaTex – empirischer Teil II.....	122
Tabelle 54: Verteilung der Gesamtpunktwerte nach Schulabschluss – empirischer Teil II.....	123
Tabelle 55: Allgemeine Daten zur Rasch-Analyse - empirischer Teil II ..	124
Tabelle 56: Überblick Rasch-Analyse – empirischer Teil II.....	125

Tabelle 57: RUMM 2030: Überblick Rasch-Analyse - empirischer Teil II.....	125
Tabelle 58: RUMM 2030: Übersicht Item-Person und Item-Trait Interaction - empirischer Teil II.....	127
Tabelle 59: RUMM 2030: Übersicht Individual Item-Fit Werte - empirischer Teil II.....	127
Tabelle 60: RUMM 2030: Ergebnisse χ^2 basierende Itemreduktion - empirischer Teil II.....	128
Tabelle 61: WINMIRA: Übersicht Personenparameter - empirischer Teil II.....	128
Tabelle 62: WINMIRA: Übersicht Mittelwert Personenparameterschätzung - empirischer Teil II	129
Tabelle 63: WINMIRA: Übersicht Itemschwierigkeiten der Items - empirischer Teil II.....	129
Tabelle 64: WINMIRA: Übersicht Q-Indizes und ZQ-Werte - empirischer Teil II.....	130
Tabelle 65: WINMIRA: Übersicht Ergebnisse Modellgeltungstest - empirischer Teil II.....	130
Tabelle 66: WINMIRA: Übersicht Ergebnisse Bootstrap - empirischer Teil II	131
Tabelle 67: eRm: Übersicht Itemschwierigkeiten - empirischer Teil II	132
Tabelle 68: eRm: Ergebnisse Modellgeltungstest - empirischer Teil II ...	132
Tabelle 69: eRm: Ergebnisse Wald-Test - empirischer Teil II.....	134
Tabelle 70: eRm: Übersicht Personenparameter - empirischer Teil II	135
Tabelle 71: Raum- und Inventarplan des Altenpflegeheims St. Josef (Erdgeschoss und 1. Obergeschoss).....	138
Tabelle 72: Entfernungen deutscher Städte/Dörfer in km (Landstraßen)	141
Tabelle 73: Auszug aus der Zutatenliste für Cocktails.....	142
Tabelle 74: Lösungswege zu den Aufgaben 6 - 8	144
Tabelle 75: Prozentrangnormen Schulabschlüsse (nach Jordan 2011a, S. 357)	147
Tabelle 76: Übersicht Leistungsbeurteilung (nach Schneider et al. 2007, S. 20)	148
Tabelle 77: WINMIRA: Schülerleistungen in Bezug auf den TeMaTex ..	149
Tabelle 78: Beschreibung der Kompetenzstufen (modifiziert nach Jordan 2011a, S. 351)	150
Tabelle 79: WINMIRA: Kompetenzstufen in Bezug auf den TeMaTex ...	151
Tabelle 80: Übersicht Kompetenzniveau sortiert nach Schulabschlüssen – Teil 1	153
Tabelle 81: RUMM 2030: Kompetenzstufen in Bezug auf den TeMaTex.....	153
Tabelle 82: RUMM 2030: Schülerleistungen in Bezug auf den TeMaTex.....	154
Tabelle 83: Übersicht Kompetenzniveau sortiert nach Schulabschlüssen - Teil 2	155
Tabelle 84: eRm: Kompetenzstufen in Bezug auf den TeMaTex	155

Tabelle 85: Übersicht Kompetenzniveau sortiert nach Schulabschlüssen - Teil 3	156
Tabelle 86: WINMIRA: Kompetenzstufen in Bezug auf den CareMaTex	158
Tabelle 87: WINMIRA: Schülerfähigkeit in Bezug auf den CareMaTex nach der Itemelimination	159
Tabelle 88: Übersicht über das nach Schulabschlüssen sortierte Kompetenzniveau – Teil 4	160
Tabelle 89: WINMIRA: Kompetenzstufen in Bezug auf den CareMaTex nach Itemelimination	160
Tabelle 90: RUMM 2030: Kompetenzstufen in Bezug auf den CareMaTex	161
Tabelle 91: RUMM 2030: Schülerfähigkeit in Bezug auf den CareMaTex	162
Tabelle 92: Übersicht Kompetenzniveau sortiert nach Schulabschlüssen - Teil 5	163
Tabelle 93: eRm: Kompetenzstufen in Bezug auf den CareMaTex	163
Tabelle 94: eRm: Schülerfähigkeiten in Bezug auf den CareMaTex	164
Tabelle 95: Übersicht Kompetenzniveau sortiert nach Schulabschlüssen - Teil 6	165

17 Abbildungsverzeichnis

Abbildung 1: Aufgabe zur Prüfungsvorbereitung – Fallbeispiel (nach Derrer-Merk 2013, S. 156 f.).....	32
Abbildung 2: Aufgabe zur Prüfungsvorbereitung - Aufgabenstellung (nach Derrer-Merk 2013, S. 161).....	33
Abbildung 3: Situation-Problem-Solver-Modell (nach Reusser 1992, S. 231).....	44
Abbildung 4: eigene Darstellung der Item Characteristic Curve 1-pl Rasch-Modell.....	68
Abbildung 5: Statistische Darstellung der Kompetenzstufen des TeMaTex (nach Jordan 2011a, S. 350).....	81
Abbildung 6: Aufgabentext 1 des TeMaTex (nach Jordan & Stein 2011c, S. 2).....	83
Abbildung 7: Item 1 zum Aufgabentext 1 (nach Jordan & Stein 2011c, S. 4).....	85
Abbildung 8: Item 2 zum Aufgabentext 1 (nach Jordan & Stein 2011c, S. 4).....	85
Abbildung 9: RUMM 2030: Verteilung Personenfähigkeit und Itemschwierigkeit - empirischer Teil I.....	98
Abbildung 10: RUMM 2030: Item Characteristic Curve von Item 10 - empirischer Teil I.....	100
Abbildung 11: RUMM 2030: Item Characteristic Curve von Item 13 - empirischer Teil I.....	101
Abbildung 12: RUMM 2030: Item Characteristic Curve von Item 17 - empirischer Teil I.....	101
Abbildung 13: eRm: Ergebnis grafischer Modellgeltungstest - empirischer Teil I.....	111
Abbildung 14: eRm: Item Characteristic Curves der 18 Items - empirischer Teil I.....	112
Abbildung 15: eRm: Person-Item-Map - empirischer Teil I.....	113
Abbildung 16: RUMM 2030: Verteilung Personenfähigkeit und Itemschwierigkeit – empirischer Teil II.....	124
Abbildung 17: RUMM 2030: Item Characteristic Curve von Item 14 - empirischer Teil II.....	126
Abbildung 18: RUMM 2030: Item Characteristic Curve von Item 16 - empirischer Teil II.....	126
Abbildung 19: eRm: Ergebnis grafischer Modellgeltungstest - empirischer Teil II.....	133
Abbildung 20: eRm: Item Characteristic Curves der 17 Items - empirischer Teil II.....	134
Abbildung 21: eRm: Person-Item-Map - empirischer Teil II.....	135
Abbildung 22: Aufgabentext 1: Renovierung eines Altenpflegeheimes ..	137
Abbildung 23: Kraftstoffverbrauch in L (pro 100 km) in Abhängigkeit der Durchschnittsgeschwindigkeit.....	141
Abbildung 24: Anteil an reinem Alkohol in verschiedenen alkoholischen Getränken.....	142

18 Anhang

Anhang 1: Anschreiben und Erhebungsbogen zum Expertenrating

Anhang 2: Version des CareMaTex für das Expertenrating

Anhang 3: Hinweise zur Durchführung

Anhang 4: R-Skript

Anhang 5: Rechenschritte zur Berechnung der Differenz zwischen Personen- und Itemparametern

Anhang 6: Versicherung selbständiger Arbeit

Anhang 1:

Expertenbefragung/Expertenrating

Sehr geehrte Damen und Herren,

die Überarbeitung bzw. die Anpassung der Testaufgaben des TeMaTex auf den Kontext der Altenpflege geht in die letzte Phase. Nun ist es wichtig, dass die überarbeiteten Testaufgaben von Experten aus dem pädagogischen Bereich der Altenpflege nach ihrer Relevanz zur Erfassung von mathematischen Fähigkeiten beurteilt werden - auch Expertenrating genannt.

Wie genau können Sie uns dabei helfen? ...

Erster Schritt:

Im Anhang finden Sie die überarbeiteten Testaufgaben inklusive Lösungswege. Lesen Sie sich die Testaufgaben sorgfältig durch. Gerne können Sie die Testaufgaben auch bearbeiten und Ihre Lösungen mit den Lösungen in der Datei vergleichen.

Zweiter Schritt:

Nach der Sichtung der Testaufgaben füllen Sie bitte den beiliegenden Evaluationsbogen aus.

Ich bedanke mich im Voraus für Ihre Mitarbeit und verbleibe mit freundlichen Grüßen

Thomas Knappich, MScN
Projektleiter

Bitte senden Sie den Evaluationsbogen mit dem beiliegenden frankierten Umschlag so bald wie möglich zurück an folgende Adresse:

Thomas Knappich
Ettlinger Str. 21/2
76448 Durmersheim

Evaluationsbogen Expertenrating					
Name des Experten:					
Evaluationsfrage:	sehr relevant	relevant	wenig relevant	nicht relevant	Kommentar:
Ich halte die zur Lösung von Aufgabe 1 benötigten Fähigkeiten für den Altenpflegeberuf für...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Ich halte die zur Lösung von Aufgabe 2 benötigten Fähigkeiten für den Altenpflegeberuf für...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Ich halte die zur Lösung von Aufgabe 3 benötigten Fähigkeiten für den Altenpflegeberuf für...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Ich halte die zur Lösung von Aufgabe 4 benötigten Fähigkeiten für den Altenpflegeberuf für...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Ich halte die zur Lösung von Aufgabe 5 benötigten Fähigkeiten für den Altenpflegeberuf für...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Ich halte die zur Lösung von Aufgabe 6 benötigten Fähigkeiten für den Altenpflegeberuf für...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

Ich halte die zur Lösung von Aufgabe 7 benötigten Fähigkeiten für den Altenpflegeberuf für...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Ich halte die zur Lösung von Aufgabe 8 benötigten Fähigkeiten für den Altenpflegeberuf für...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Evaluationsfrage:	sehr relevant	relevant	wenig relevant	nicht relevant	Kommentar:
Ich halte die zur Lösung von Aufgabe 9 benötigten Fähigkeiten für den Altenpflegeberuf für...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Ich halte die zur Lösung von Aufgabe 10 benötigten Fähigkeiten für den Altenpflegeberuf für...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Ich halte die zur Lösung von Aufgabe 11 benötigten Fähigkeiten für den Altenpflegeberuf für...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Ich halte die zur Lösung von Aufgabe 12 benötigten Fähigkeiten für den Altenpflegeberuf für...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Ich halte die zur Lösung von Aufgabe 13 benötigten Fähigkeiten für den Altenpflegeberuf für...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Ich halte die zur Lösung von Aufgabe 14 benötigten Fähigkeiten für den Altenpflegeberuf für...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Ich halte die zur Lösung von Aufgabe 15 benötigten Fähigkeiten für den Altenpflegeberuf für...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Ich halte die zur Lösung von Aufgabe 16 benötigten Fähigkeiten für den Altenpflegeberuf für...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Ich halte die zur Lösung von Aufgabe 17 benötigten Fähigkeiten für den Altenpflegeberuf für...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

Evaluationsfrage:	trifft zu	trifft eher zu	trifft eher nicht zu	trifft nicht zu	Kommentar:
Der Test in seiner Gesamtheit gibt das Anforderungsprofil im Bereich der Fähigkeit „Verständnis mathemathikhaltiger Texte“ für den Altenpflegeberuf angemessen wieder:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

Zusätzliche Kommentare und Anregungen:
<div></div>

Anhang 2:

Mathematik für Altenpflegeschüler

Sehr geehrte Schülerinnen und Schüler,

im Rahmen eines Forschungsprojektes an der Philosophisch-Theologischen Hochschule in Vallendar wird, in Zusammenarbeit mit verschiedenen Pflegeschulen, an der Entwicklung eines neuen Testverfahrens gearbeitet. Mithilfe des Testverfahrens soll ermittelt werden, welche mathematischen Kenntnisse Schüler- und Schülerinnen zu Beginn und während der Berufsausbildung haben, um perspektivisch Fördermaterialien in der Ausbildung zur Altenpflege anbieten zu können.

Um Informationen über den Leistungsstand von Altenpflegeschülern bzw. angehenden Altenpflegeschülern zu erhalten, bitten wir Sie an dem folgenden Test teilzunehmen. Beachten Sie dabei die folgenden Durchführungshinweise.

Durchführungshinweise:

- (1) Die Bearbeitungszeit für den Test beträgt 45 Minuten (inklusive der Testinstruktion). Es sind **keine Hilfsmittel (Taschenrechner, Bücher, etc.)** erlaubt.
- (2) Die Lösungen zu den Aufgaben tragen Sie bitte in die dafür vorgesehenen fett markierten Felder ein. Nebenrechnungen können Sie auf den karierten Flächen durchführen.

Um Ihre Ergebnisse einordnen zu können, bitten wir Sie, dass Sie die unten stehende Tabelle vervollständigen.

Mit freundlichen Grüßen

Thomas Knappich, MScN
Projektleiter

Bitte geben Sie uns für die Auswertung der Tests folgende persönliche Informationen:

Datum der Testdurchführung: _____

Name: _____

Klasse: _____

Lehrer: _____

Geschlecht: ☐ Männlich ☐ Weiblich

Schulabschluss: ☐ Kein Schulabschluss ☐ Hauptschulabschluss

☐ Realschulabschluss ☐ Fachabitur

☐ Abitur

Text 1: Renovierung eines Altenpflegeheimes

Für die Renovierung des Altenpflegeheims „St. Josef“ ist die Firma Renofix GmbH zuständig. Im Jahr 2015 standen im Rahmen der Umsetzung der Heimmindestbauverordnung einige Renovierungsarbeiten an. Der Einrichtungsträger Caritas hat mit der Firma Renofix GmbH einen Vertrag abgeschlossen, der die Kosten bzgl. der Renovierungsarbeiten regelt (vgl. Vertragsauszug unten).

Bearbeiten Sie die Aufgaben zu diesem Text. Beachten Sie dazu den folgenden Vertragsauszug und die Tabelle auf der folgenden Seite.

Vertragsauszug zwischen der Firma Renofix GmbH und des Einrichtungsträgers Caritas:

§ 5: Kosten für Renovierungsarbeiten:

(1) Die Bezahlung für durchgeführte Renovierungsarbeiten richtet sich nicht nach der Arbeitszeit, sondern wird anhand des Materialverbrauchs berechnet. Demnach stellt die Firma Renofix GmbH dem Einrichtungsträger folgende Beträge in Rechnung:

- ☛ pro m² rutschfester Teppich: 15 €
- ☛ pro m verlegter Bodenleisten: 6,50 €
- ☛ pro m² gestrichener Wand: 1,30 €.

In diesen Beträgen sind die Arbeitskosten bereits enthalten.

(2) Die Firma Renofix GmbH erstellt die Rechnungen für Renovierungsarbeiten auf Grundlage der Tabelle „Raum- und Inventarplan des Altenpflegeheims St. Josef“. Bei den zu streichenden Wandflächen werden Türen und Fensterflächen nicht einzeln berechnet, sondern es wird pauschal 20 % der gesamten Wandfläche abgezogen. Die Decke wird nicht gestrichen.

Tabelle: Raum- und Inventarplan des Altenpflegeheims St. Josef (Erdgeschoss und 1. Obergeschoss)

	Erdgeschoss										1. Obergeschoss									
	Raum 0.01	Raum 0.02	Raum 0.03	Raum 0.04	Raum 0.05	Raum 0.06	Raum 0.07	Raum 0.08	Raum 0.09	Raum 0.10	Raum 1.01	Raum 1.02	Raum 1.03	Raum 1.04	Raum 1.05	Raum 1.06	Raum 1.07	Raum 1.08	Raum 1.09	Raum 1.10
Angaben zum Raum (alle Räume eine rechteckige Grundfläche)																				
Länge (in m)	8	6,5	7	4,7	12	15	6	8	5	4	7	4	5	20	5	3	6	12	5	7
Breite (in m)	6	4	5,5	5	12	20	7	4	3	6	5	9	6	15	5	6	4	17	6	5
Höhe (in m)	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
Anzahl der Türen	1	1	2	2	1	2	1	2	1	1	1	1	1	2	1	2	1	3	1	1
Breite pro Tür (in m)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Angaben zum Inventar im Raum																				
Anzahl Regale	4	3	3	5	4	7	4	2	2	3	4	3	4	8	3	2	1	4	2	2
Anzahl Schränke	2	1	2	2	4	5	3	2	2	1	2	3	1	4	2	3	1	3	2	1
Anzahl Stühle	5	5	7	2	7	16	6	4	3	2	4	6	8	15	3	6	5	16	6	8
Anzahl Tische	2	2	4	2	7	12	3	3	2	2	4	3	5	14	3	4	3	11	3	5
Anzahl Computer	3	2	5	2	10	14	5	2	1	0	3	4	3	10	5	3	2	8	6	5
Anzahl Monitore	3	2	10	4	16	14	5	3	1	0	4	5	3	10	6	3	2	10	6	7

Aufgaben zum Text 1: „Renovierung eines Altenpflegeheimes“:

Sie dürfen zur Bearbeitung der Aufgaben KEINEN Taschenrechner benutzen! Notieren Sie Ihre Rechnung auf der rechten Seite und tragen Sie das Ergebnis in das fett markierte Feld ein.

1. Aufgabe:

Der Raum 0.06 wird komplett neu renoviert. Berechnen Sie, wie viel m² rutschfester Bodenbelag verlegt werden müssen.

Antwort zu Aufgabe 1:

m²

2. Aufgabe:

Berechnen Sie, wie viel m Bodenleisten im Raum 0.06 verlegt werden müssen. Beachten Sie dabei, dass die Breite der Türen abgezogen werden muss (vgl. Tabelle)!

Antwort zu Aufgabe 2:

m

3. Aufgabe:

Berechnen Sie, wie groß die Fläche ist, die von der Firma Renofix GmbH für das Streichen der Wände in Raum 0.06 in Rechnung gestellt wird. Beachten Sie dabei den Abschnitt (2) des Vertragsauszuges!

Antwort zu Aufgabe 3:

m²

Aufgaben zum Text 1: „Renovierung eines Altenpflegeheimes“:

Sie dürfen zur Bearbeitung der Aufgaben **KEINEN Taschenrechner** benutzen! Notieren Sie Ihre Rechnung auf der rechten Seite und tragen Sie das Ergebnis in das fett markierte Feld ein.

4. Aufgabe:

Der Raum 2.04 (im 2. Obergeschoss) wird ebenfalls komplett renoviert. Er hat eine Grundfläche von 600 m². Berechnen Sie die Kosten für die Verlegung des rutschfesten Bodenbelages.

Antwort zu Aufgabe 4:

€

5. Aufgabe:

In dem Raum 2.04 müssen insgesamt 58 m an Bodenleisten verlegt werden (nach Abzug der Türbreiten). Berechnen Sie, wie teuer die Verlegung der Bodenleisten ist.

Antwort zu Aufgabe 5:

€

6. Aufgabe:

In dem Raum 2.04 wird nach der Berechnungsvorschrift (2) aus dem Arbeitsvertrag mit einer Wandfläche von 176 m² kalkuliert. Wie teuer ist die Streichung des Raumes?

Antwort zu Aufgabe 6:

€

Text 2: Der ambulante Pflegedienst

Der Altenpfleger Thorsten Müller, der in Ettlingen wohnt, ist jeden Tag in der Woche mit dem Dienstauto (Opel, 60 PS, Diesel-Motor) unterwegs. In der 23. Kalenderwoche im Jahr 2015 hatte Herr Müller an einem Tag eine besonders stressige Tour:

Zu Beginn seines Dienstes um 06:30 Uhr fährt Herr Müller von seiner Dienststelle in Ettlingen los nach Reichenbach, da er dort bei einem älteren Herrn den Blutzucker messen und Insulin spritzen muss. Im Anschluss fährt er von Reichenbach nach Langensteinbach weiter, da er dort bei einem weiteren Kunden eine kleine Toilette durchzuführen hat. In Langensteinbach hat Herr Müller noch zwei weitere Kunden zu betreuen und bleibt dort bis 10:30 Uhr. Gegen 11:00 Uhr macht sich Herr Müller wieder auf den Weg zu seiner Dienststelle. Er muss auf dem Weg zurück zur Dienststelle allerdings noch einen kurzen Hausarztbesuch in Ittersbach wahrnehmen, um ein Rezept für eine Klientin abzuholen. Zwischen 12:00 Uhr und 13:00 Uhr muss Herr Müller erst noch nach Grünwettersbach und anschließend nach Kleinsteinbach, um dort Klienten zu betreuen.

Sie dürfen zur Bearbeitung der Aufgaben KEINEN Taschenrechner benutzen! Notieren Sie Ihre Rechnung auf der rechten Seite und tragen Sie das Ergebnis in das fett markierte Feld ein.

Aufgaben zum Text 2:

Verwenden Sie zur Bearbeitung der folgenden Aufgaben die Tabelle und Abbildung der folgenden Seite.

7. Aufgabe:

Wie viele km hat Herr Müller zwischen 06:30 - 10:30 Uhr mit seinem Auto zurückgelegt?

Antwort zu Aufgabe 7:

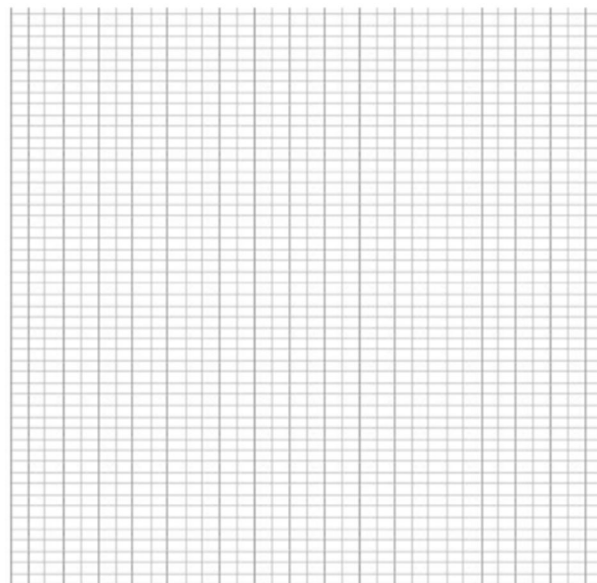
km

8. Aufgabe:

Wie viele km hat Herr Müller zwischen 11:00 – 12:00 Uhr mit seinem Auto zurückgelegt?

Antwort zu Aufgabe 8:

km



Herr Müller fährt im Durchschnitt 100 km/h. Wie viel Liter Kraftfahrtstoff hat Herr Müller auf dem Weg zwischen Grünwettersbach und Kleinsteinbach verbraucht?

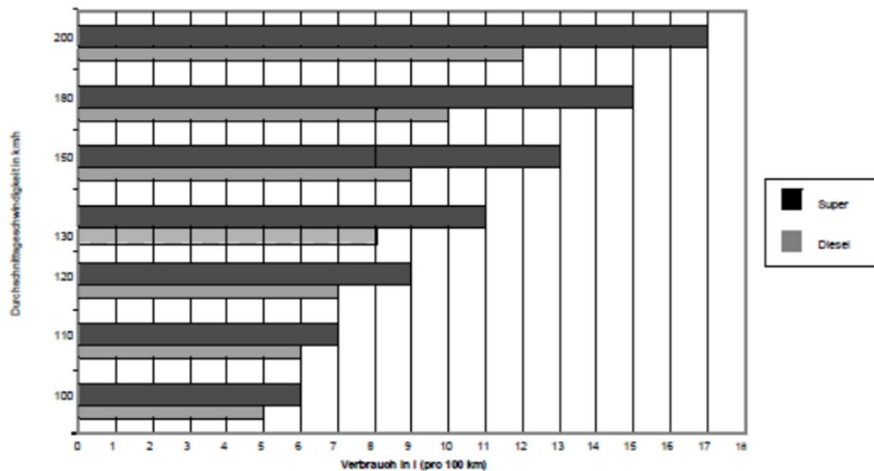
Antwort zu Aufgabe 9:

[illegible]

Tabelle 1: Entfernungen deutscher Städte/Dörfer in km (Landstraßen)

Alle Angaben in km	Langensteinbach	Reichenbach	Spielberg	Mutschelbach	Auerbach	Ettlingen	Ittersbach	Grünwettersbach	Kleinsteinbach
Langensteinbach		3,2	1,6	5,1	2,9	9,7	1,1	7,0	7,6
Reichenbach	3,2		4,8	9,0	5,5	7,0	4,2	6,0	8,4
Spielberg	1,6	4,8		7,0	4,8	12,2	0,8	8,9	9,5
Mutschelbach	5,1	9,0	7,0		7,5	17,5	6,2	9,1	2,7
Auerbach	2,9	5,5	4,8	7,5		12,0	4,0	9,4	9,9
Ettlingen	9,7	7,0	12,2	17,5	12,0		10,9	8,3	16,2
Ittersbach	1,1	4,2	0,8	6,2	4,0	10,9		8,1	8,6
Grünwettersbach	7,0	6,0	8,9	9,1	9,4	8,3	8,1		8,9
Kleinsteinbach	7,6	8,4	9,5	2,7	9,9	16,2	8,6	8,9	

Abbildung 1: Kraftfahrstoffverbrauch in L (pro 100 km) in Abhängigkeit der Durchschnittsgeschwindigkeit



Text 3: Diabetes und Alkohol

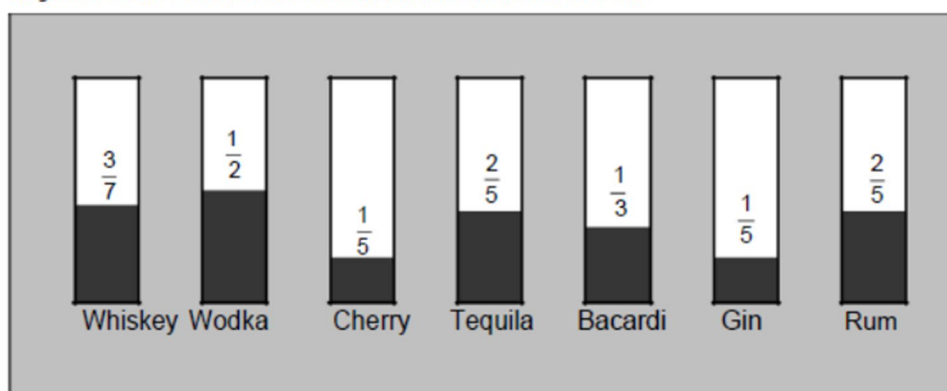
In einer Cocktailbar werden Cocktails in den Größen klein (300 ml), mittel (400 ml) und groß (500 ml) verkauft. An einem Abend erfährt Frau Maier, welche seit einem Jahr an einem Diabetes mellitus Typ II leidet, nach einem längeren Gespräch mit dem Barkeeper die Zutaten für ihren Lieblingscocktail „Mai Tai“. Der Barkeeper zeigt ihr dazu die Zutatenliste für die Cocktails (ein Auszug davon ist in der Tabelle abgebildet). Frau Maier muss aufgrund ihrer Diabetes mellitus Erkrankung darauf achten, dass Sie nicht zu viel Alkohol zu sich nimmt, damit Sie nicht in den Unterzucker gerät. Daher möchte Frau Maier nun herausfinden, wie viel Alkohol in einem „Mai Tai“ enthalten ist. Im Internet findet sie dazu das unten abgebildete Diagramm, das den Anteil an reinem Alkohol in verschiedenen alkoholischen Getränken angibt.

Bearbeiten Sie die Aufgaben auf der folgenden Seite.

Tabelle: Auszug aus der Zutatenliste für Cocktails

Zutatenverhältnis der Cocktails							
Louisiana		Long Island Ice Tea		Mai Tai		Tequila Sunrise	
Orangensaft	1/6	Cola	2/5	Rum	1/6	Orangensaft	5/8
Pfirsichnektar	1/6	Gin	1/15	Gin	1/4	Tequila	1/6
Whiskey	2/5	Orangensaft	2/15	Zitronensaft	1/6	Zitronensaft	1/8
Gin	1/6	Rum	1/10	Ananassaft	1/3	Grenadine	1/12
Zitronensaft	1/10	Tequila	2/15	Pfirsichsaft	1/20		
		Wodka	1/10	Cola	1/30		
		Zitronensaft	1/15				

Diagramm: Anteil an reinem Alkohol in verschiedenen alkoholischen Getränken



Aufgaben zum Text 3:

Sie dürfen zur Bearbeitung der Aufgaben **KEINEN Taschenrechner** benutzen! Notieren Sie Ihre Rechnung auf der rechten Seite und tragen Sie das Ergebnis in das fett markierte Feld ein.

10. Aufgabe:

Frau Maier hat an dem Abend in der Cocktailbar zwei kleine Mai Tai getrunken. Wieviel ml sind das?

Antwort zu Aufgabe 10:

ml

11. Aufgabe:

Zu welchem Anteil besteht ein Mai Tai aus Fruchtsaft?

(Geben Sie das Ergebnis als Bruch an)

Antwort zu Aufgabe 11:

(Bruch)

12. Aufgabe:

Geben Sie den Alkoholgehalt von Rum in % an.

Antwort zu Aufgabe 12:

%

Text 4: Die Reinigungsfirma

Die Reinigungsfirma Cleanfix GmbH betreibt zwei weit voneinander entfernte Zweigstellen. Insgesamt arbeiten 8 Reinigungsfachkräfte in der Firma. Die Hauptaufgabe der Reinigungsfirma besteht in der hygienischen Säuberung der Böden in den Räumlichkeiten von Pflegeeinrichtungen. Auf diese Weise kann die Reinigungsfirma mit den Mitarbeitern einer Zweigstelle pro Tag 200 m² Böden reinigen. Mit dem derzeitigen Lagerbestand an speziellen Reinigungsmitteln kann die Firma noch insgesamt für 12 Arbeitstage pro Zweigstelle die Böden der Pflegeeinrichtungen reinigen. Kurzfristig erhält die Reinigungsfirma zusätzlich den Auftrag, in beiden Pflegeeinrichtungen auch die Büroböden zu reinigen. So müssen in der ersten Pflegeeinrichtung insgesamt 300 m² Böden und in der zweiten Pflegeeinrichtung 400 m² Böden pro Tag gereinigt werden.

Aufgaben zum Text 4:

Sie dürfen zur Bearbeitung der Aufgabe **KEINEN Taschenrechner** benutzen! Notieren Sie Ihre Rechnung auf der rechten Seite und tragen Sie das Ergebnis in das fett markierte Feld ein.

13. Aufgabe:

Nach wie vielen Arbeitstagen ist der Lagerbestand an der zweiten Zweigstelle aufgebraucht?

Antwort zu Aufgabe 13:

Nach Tagen.

Aufgabentext zu Aufgabe 14 und 15:

Herr Weinfeld unterzog sich vor einer Woche einer Hüftoperation. Nach einer erfolgreichen Genesung konnte er planmäßig entlassen werden. Herr Weinfeld fährt nach seinem Krankenhausaufenthalt mit einem Taxi vom Klinikum Karlsruhe zu seiner Wohnung. Die Taxifahrt kostet 3 € zuzüglich 1,50 € pro gefahrenen km. Als Herr Weinfeld an seiner Wohnung ankommt, verlangt der Taxifahrer einen Preis von 15,75 €.

Sie dürfen zur Bearbeitung der Aufgaben KEINEN Taschenrechner benutzen! Notieren Sie Ihre Rechnung auf der rechten Seite und tragen Sie das Ergebnis in das fett markierte Feld ein.

14. Aufgabe:

Um wie viel € ist eine Taxifahrt der Entfernung von 10 km teurer als eine Taxifahrt der Entfernung 9 km?

Antwort zu Aufgabe 14:

€

15. Aufgabe:

Wie viele km wohnt Herr Weinfeld vom Klinikum entfernt?

Antwort zu Aufgabe 15:

km

Aufgabentext zu Aufgabe 16:

Die Altenpflegefachkraft Frau Kiemig benötigt zur Versorgung von Wunden eines Bewohners diverses Verbandsmaterial. Dafür geht sie ins hauseigene Materiallager. Der Mitarbeiter im Materiallager benötigt eine Leiter, um an das Material auf dem Regal zu gelangen. Damit die Leiter nicht umkippt, wird das eine Ende der Leiter 4 m von der Wand entfernt aufgestellt. Das andere Ende der Leiter erreicht so genau die Kante des Regals, das sich 3 m über dem Boden befindet.

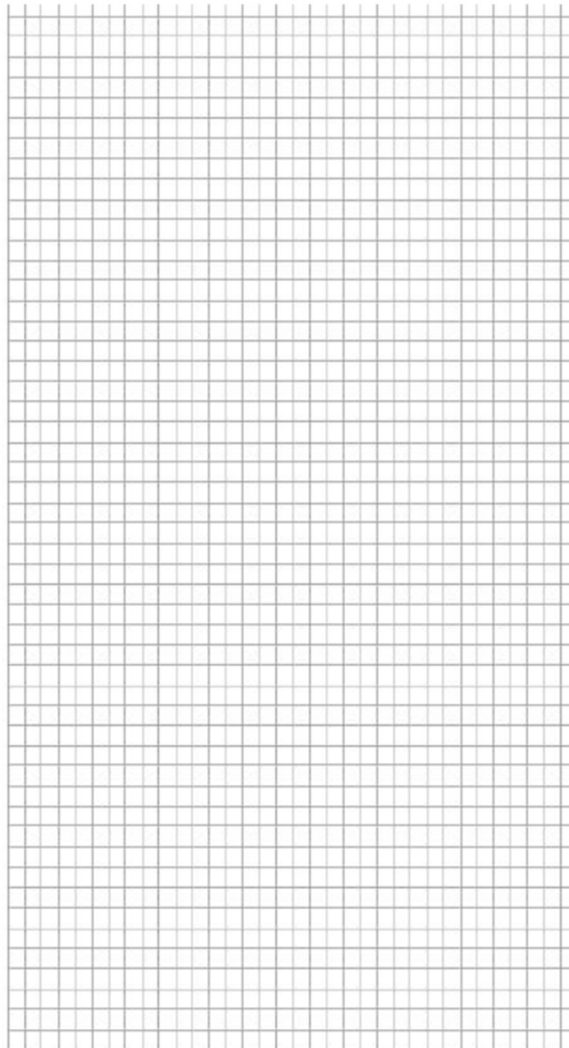
Sie dürfen zur Bearbeitung der Aufgaben KEINEN Taschenrechner benutzen! Notieren Sie Ihre Rechnung auf der rechten Seite und tragen Sie das Ergebnis in das fett markierte Feld ein.

16. Aufgabe:

Wie lang ist die Leiter?

Antwort zu Aufgabe 16:

m



Aufgabentext zu Aufgabe 17:

In einer Firma für Medizinprodukte werden verschiedene Produkte hergestellt. Unter anderem Kompressen, Mullbinden und Wundpflaster. Damit für die Kunden immer genügend Materialien zur Verfügung stehen, hat sich die Firma entschieden ein kleines Lager zu halten, um die Überproduktion kurzfristig lagern zu können. Am Anfang der 34. Kalenderwoche (KW) (also Montagmorgen) sind in diesem Lager 323 Packungen mit Wundpflaster, 409 Packungen mit Mullbinden sowie 89 Packungen mit Kompressen. Gegen Ende der 34. KW (also Freitag Abend) werden die Produkte vom Transportunternehmen „Maier-Transporte“ abgeholt und zu den Kunden geschickt – je nach Auftragslage werden etwa 100 Packungen mit Kompressen, 350 Packungen mit Wundpflaster und 500 Packungen mit Mullbinden pro Woche abgeholt. Bevor die Firma Maier Transporte am Freitagnachmittag die Produkte abholt, hat das Lager der Fabrik einen Bestand von 634 Packungen mit Mullbinden, 169 Packungen mit Kompressen und 526 Packungen mit Wundpflaster.

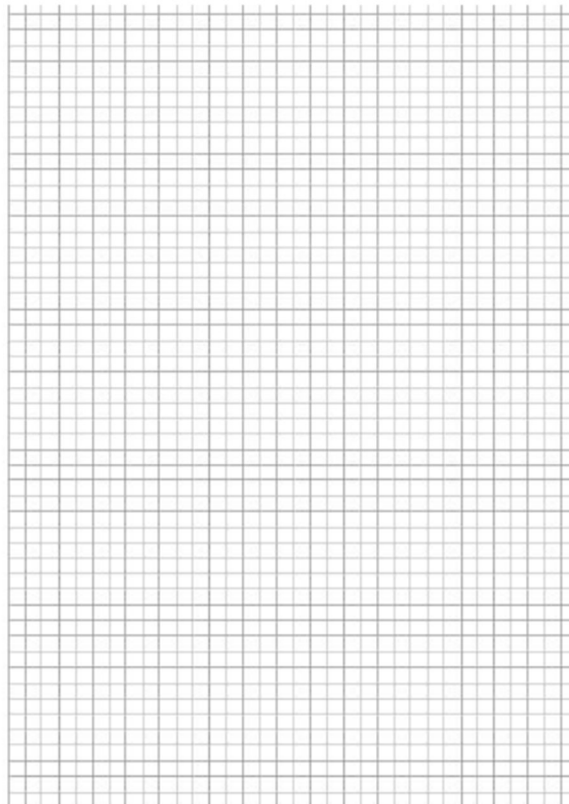
Sie dürfen zur Bearbeitung der Aufgaben **KEINEN Taschenrechner** benutzen! Notieren Sie Ihre Rechnung auf der rechten Seite und tragen Sie das Ergebnis in das fett markierte Feld ein.

17. Aufgabe:

Wie viele Packungen mit Mullbinden wurden in der 34. KW hergestellt?

Antwort zu Aufgabe 17:

--



**Aufgaben und Lösungen für die überarbeitete Test-
version**

Aufgabe	1	2	3	4	5
Korrektes Ergebnis	300	68	168	900	377

Aufgabe	6	7	8	9	10
Korrektes Ergebnis	228,8	10,2	12	0,445	600

Aufgabe	11	12	13	14	15
Korrektes Ergebnis	11/20	40	6	1,5	8,5

Aufgabe	16	17	
Korrektes Ergebnis	5	225	

Anhang 3:

Testanleitung

Anwendungsbereich:

Der CareMaTex kann in Altenpflegeschulen zur Diagnose von Schwächen im Bereich des mathematischen Textverständnisses als Gruppen- oder Einzeltest eingesetzt werden.

Durchführungsbestimmungen:

Die Durchführungszeit des Tests inklusive Instruktionen dauert 45 Minuten. Die Durchführung kann in einem Klassenraum stattfinden. Bei der Auswahl des Raumes sollte beachtet werden, dass die Lärmbelästigung von außen möglichst gering ausfällt. Hilfsmittel wie z.B. Taschenrechner oder Bücher sind bei der Bearbeitung der Testaufgaben nicht erlaubt. Um Abschreiben bei der Durchführung als Gruppentest zu verhindern, sollten die Probanden möglichst weit auseinander gesetzt werden. Für das Bearbeiten des Tests benötigt jeder Proband folgende Materialien:

1. 1 Test
2. 1 Stift

Der Test beginnt mit der Instruktion durch den Testleiter und endet nach 45 Minuten (inklusive Instruktionszeit). Während der Bearbeitungszeit – also nach Abschluss der Testinstruktion – werden vom Testleiter keine Fragen beantwortet.

Wörtliche Instruktion:

Zunächst achtet der Testleiter darauf, dass alle Probanden den Test und einen Stift vor sich haben, damit die Bearbeitung beginnen kann. Der Testleiter hat ebenfalls einen Test zu Präsentationszwecken bereitlegt. Er beginnt anschließend mit der Instruktion (Der Sprechtext ist kursiv gedruckt):

Sie bearbeiten heute einen Test mit mathematischen Aufgaben. Dazu haben Sie einen Test mit den Aufgabentexten und Aufgaben bekommen. Sie benötigen zudem einen Stift zum Lösen der Aufgaben.

Der Testleiter zeigt dabei den Test. Der Testleiter weist daraufhin, dass der Test doppelseitig gedruckt wurde!

Der Test wurde doppelseitig gedruckt. Bitte füllen Sie zunächst den unteren Teil im Anschreiben des Tests aus.

Der Testleiter wartet, bis alle den unteren Teil des Anschreibens des Tests ausgefüllt haben.

Im Anschreiben des Tests finden Sie auch die Bearbeitungshinweise. Diese lese ich jetzt vor: Die Bearbeitungszeit für den Test beträgt 45 Minuten (inklusive der Testinstruktion). Es sind keine Hilfsmittel (Taschenrechner, Bücher, etc.) erlaubt. Die Lösungen zu den Aufgaben tragen Sie bitte in die

dafür vorgesehenen fett markierten Felder ein. Nebenrechnungen können Sie auf den karierten Flächen durchführen. Der Test ist wie folgt aufgebaut: Schlagen Sie dazu die Seite 2 des Tests auf.

Der Testleiter wartet, bis alle Probanden die zweite Seite aufgeschlagen haben.

Zunächst gibt es immer einen Aufgabentext - teilweise mit weiteren Materialien wie Tabellen oder Abbildungen. Auf dieser Seite sehen Sie den Aufgabentext 1: Renovierung eines Altenpflegeheimes. Auf der nächsten Seite ist noch eine Tabelle zu dem Aufgabentext abgebildet. Blättern Sie nun um auf Seite 4.

Der Testleiter wartet, bis alle Probanden die vierte Seite aufgeschlagen haben.

Hier sehen Sie nun die Aufgaben zu dem Aufgabentext 1: Renovierung eines Altenpflegeheimes. Diese müssen Sie nun bearbeiten. Nebenrechnungen können Sie auf den karierten Flächen durchführen. Die Lösungen zu den Aufgaben tragen Sie bitte in die dafür vorgesehenen fett markierten Felder ein. Fahren Sie mit allen der insgesamt 7 Aufgabentexte und 17 Aufgaben so fort. Wenn Sie mit der Bearbeitung aller Aufgaben fertig sind, schließen Sie den Test. Die Tests werden am Ende der Bearbeitungszeit eingesammelt.

Der Testleiter wartet bis zum Ablauf der Bearbeitungszeit und sammelt die Tests ein.

Anhang 4:

R-Skript

Bedingte Maximum-Likelihood-Schätzung der Aufgaben-Parameter

Laden des eRm-Paketes:

```
> library("eRm")
```

Installation des eRm packages:

```
> install.packages("eRm",dependencies=TRUE)
```

Anpassen eines Rasch-Modells und Ausgabe der Aufgaben-Parameter:

```
> rm1<-RM(Datenmatrix_Rohwerte_040814...Kopie)
> rm1$betapar und
> summary(rm1)
```

Modellkontrolle und Aufgaben-Selektion

Überprüfung des Modells mithilfe von Andersens Likelihood-Quotienten-Test. Als Splitkriterium wird der Mittelwert der Personen-Randsummen verwendet:

```
> lr1 <- LRtest(rm1, splitcr="mean", se=TRUE)
> lr1
```

Überprüfung der Aufgaben in Bezug auf die Modellverletzung anhand des grafischen Modelltests mit 95 %-Konfidenz-Regionen. Als Splitkriterium wird wieder der Mittelwert der Personen-Randsummen verwendet:

```
> plotGOF(lr1, xlab="Randsumme < Mittelwert",
+ ylab="Randsumme >= Mittelwert", tlab="number",
+ conf=list(gamma=0.95, col=1), main=" ")
```

Aufgaben-spezifischen Wald- Test zur Überprüfung eines signifikanten Differential Item Functioning. Als Splitkriterium wird wieder der Mittelwert der Personen-Randsummen verwendet:

```
> Waldtest(rm1, splitcr="mean")
```

Löschen der Aufgaben mit Differential Item Functioning:

```
> dat2 <- Datenmatrix_Rohwerte_040814...Kopie [,-c(7,9,13)]
```

Anpassung eines neuen Rasch-Modells nach Ausschluss der Aufgaben:


```

> rm2 <- RM (dat2)

# Nochmalige Überprüfung des Modells mithilfe von Andersens Like-
likelihood-Quotienten- Test. Als Splitkriterium wird wieder der Mittelwert der
Personen-Randsummen verwendet:

> lr2 <- LRtest(rm2, splitcr="mean", se=TRUE)
> lr2

### Grafische Darstellung der ICCs der 18 Aufgaben:

> plotjointICC (rm1)

### Person-Item-Map der Lage der Personen- und Aufgaben-Parameter
auf der latenten Skala:

> plotPImap (rm1)

### Schätzung der Personen-Parameter:

> pp2 <- person.parameter(rm1)
> pp2

## Interrater-Reliabilität/Konkordanz-Koeffizienz

# Laden des irr-Paketes:

> library("irr")

# Berechnen des Koeffizienten:

> rater1 <- c(2,2,3,2,3,2,2,2,2,2,2,1,2,2,2,3,2)
> rater2 <- c(1,1,3,1,1,1,1,1,1,1,3,3,1,1,1,3,1)
> rater3 <- c(4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4)
> rater4 <- c(2,2,3,2,2,2,1,1,2,1,2,1,3,2,4,4,2)
> rater5 <- c(3,3,3,3,0,3,2,2,2,1,1,2,2,2,2,3,3)
> rater6 <- c(3,4,3,3,4,3,2,2,2,2,2,2,3,2,2,4,2)
> rater7 <- c(2,2,3,2,2,2,1,1,2,1,2,2,2,2,2,3,2)
> rater <- cbind(rater1,rater2,rater3,rater4,rater5,rater6,rater7)
> rater
> kendall(rater)

```

Anhang 5:

$$p(x_{vi} = 1) = 0,70 = \frac{\exp(x_{vi}(\theta_{vi} - \sigma_i))}{1 + \exp(\theta_{vi} - \sigma_i)}$$

Im zweiten Schritt wird mit $1 + \exp(\theta_{vi} - \sigma_i)$ multipliziert:

$$0,70 + 0,70 \exp(\theta_{vi} - \sigma_i) = \exp(\theta_{vi} - \sigma_i)$$

Im dritten Schritt wird mit $0,70 \exp(\theta_{vi} - \sigma_i)$ subtrahiert:

$$0,70 = 0,30 \exp(\theta_{vi} - \sigma_i)$$

Im vierten Schritt wird durch 0,70 dividiert und logarithmiert:

$$\ln\left(\frac{0,70}{0,30}\right) = \theta_{vi} - \sigma_i$$

Im fünften Schritt wird σ_i addiert:

$$\ln\left(\frac{0,70}{0,30}\right) + \sigma_i = \theta_{vi}$$

Zuletzt wird die Gleichung umgestellt und berechnet:

$$\theta_{vi} = \sigma_i + \ln\left(\frac{p}{1-p}\right) =$$

$$\theta_{vi} = \sigma_i + \ln\left(\frac{0,70}{1-0,70}\right) =$$

$$\theta_{vi} = \sigma_i + \ln\left(\frac{0,70}{0,30}\right) =$$

$$\theta_{vi} = \sigma_i + 0,847$$

Durch diese Vorgehensweise können die jeweiligen Personenparameter bestimmt werden. Es ist aber auch möglich, die jeweiligen Itemparameter zu berechnen. Hierfür wird die Zahl 0,847 subtrahiert. Nach Umstellung der Gleichung entsteht somit folgendes Ergebnis:

$$\sigma_i = \theta_{vi} - 0,847$$

Anhang 6:

Versicherung selbständiger Arbeit

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe.

Ich versichere, dass die vorliegende Arbeit bisher nicht veröffentlicht wurde und nicht bereits einmal an einer anderen Hochschule als Dissertation eingereicht wurde.

Die eingereichte schriftliche Fassung entspricht der auf dem elektronischen Speichermedium (CD-Rom).

Durmersheim, 03.01.2017
